

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

# REPORTS

Ad1p sequence following Ser<sup>200</sup> and occurs within the domain of Ad1p that shows homology with HDE (14). To delete the complete STE23 sequence and create the ste23Δ:URA3 mutation, polymerase chain reaction (PCR) primers (5'-TCGGAAGACGTCAT-TCTTGCTCATTGATATTGCTC- GTAGATTG-TACTGAGAGTGCAC-3' and 5'-GCTACAAACAGC-GTGGACTTGAATGCCCCGACATCTTCTGACTGT-GCGGTATTTCACACCG-3') were used to amplify the URA3 sequence of pRS316, and the reaction product was transformed into yeast for one-step gene replacement [R. Rothstein, *Methods Enzymol.* 184, 281 (1991)]. To create the ad1Δ::LEU2 mutation contained on p114, a 5.0-kb Sal I fragment from pAX1 was cloned into pUC19, and an internal 4.0-kb Hpa I-Xho I fragment was replaced with a LEU2 fragment. To construct the ste23Δ:LEU2 allele (a deletion corresponding to 931 amino acids) carried on p153, a LEU2 fragment was used to replace the 2.6-kb Pml I-Ecl136 II fragment of STE23, which occurs within a 6.2-kb Hind III-Bgl II genomic fragment carried on pSP72 (Promega). To create YEpMFA1, a 1.8-kb Bam HI fragment containing MFA1, from pK16 [K. Kuchler, R. E. Stern, J. Thorne, *EMBO J.* 8, 3973 (1989)], was ligated into the Bam HI site of YEp351 [J. E. Hill, A. M. Myers, T. J. Koerner, A. Tzagoloff, *Yeast* 2, 163 (1986)].

uct. pC225 is a KS+ (Stratagene) plasmid containing a 0.5-kb Bam HI-Sal I fragment from pAX1. Substitution mutations of the proposed active site of Ad1p were created with the use of pC225 and site-specific mutagenesis involving appropriate synthetic oligonucleotides (ad1-H68A, 5'-GTGCTCACAAGCGCT-GCCAAACCGGC-3'; ad1-E71A, 5'-AAGAATCAT-GTGGGCACAAAGGTGGCG-3'; and ad1-E71D, 5'-AAGAATCATGTGATCACAAGGTGGCG-3'). The mutations were confirmed by sequence analysis. After mutagenesis, the 0.4-kb Bam HI-Msc I fragment from the mutagenized pC225 plasmids was transferred into pAX1 to create a set of pRS316 plasmids carrying different AX1 alleles, p124 (ad1-H68A), p130 (ad1-E71A), and p132 (ad1-E71D). Similarly, a set of HA-tagged alleles carried on YEp352 were created after replacement of the p151 Bam HI-Msc I fragment, to generate p161 (ad1-E71A), p162 (ad1-

32

N. Davis, T. Favero, C. de Hoog, and S. Kim for comments on the manuscript. Supported by a grant to C.B. from the Natural Sciences and Engineering Research Council of Canada. Support for M.N.A. was from a California Tobacco-Related Disease Research Program postdoctoral fellowship (4FT-0063).

22 June 1995; accepted 21 August 1995

## Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray

Mark Schena,\* Dari Shalon,\*† Ronald W. Davis, Patrick O. Brown‡

A high-capacity system was developed to monitor the expression of many genes in parallel. Microarrays prepared by high-speed robotic printing of complementary DNAs on glass were used for quantitative expression measurements of the corresponding genes. Because of the small format and high density of the arrays, hybridization volumes of 2 microliters could be used that enabled detection of rare transcripts in probe mixtures derived from 2 micrograms of total cellular messenger RNA. Differential expression measurements of 45 *Arabidopsis* genes were made by means of simultaneous, two-color fluorescence hybridization.

The temporal, developmental, topographical, histological, and physiological patterns in which a gene is expressed provide clues to its biological role. The large and expanding database of complementary DNA (cDNA) sequences from many organisms (1) presents the opportunity of defining these patterns at the level of the whole genome.

For these studies, we used the small flowering plant *Arabidopsis thaliana* as a model organism. *Arabidopsis* possesses many advantages for gene expression analysis, including the fact that it has the smallest genome of any higher eukaryote examined to date (2). Forty-five cloned *Arabidopsis* cDNAs (Table 1), including 14 complete sequences and 31 expressed sequence tags (ESTs), were used as gene-specific targets. We obtained the ESTs by selecting cDNA clones at random from an *Arabidopsis* cDNA library. Sequence analysis revealed that 28 of the 31 ESTs matched sequences

in the database (Table 1). Three additional cDNAs from other organisms served as controls in the experiments.

The 48 cDNAs, averaging ~1.0 kb, were amplified with the polymerase chain reaction (PCR) and deposited into individual wells of a 96-well microtiter plate. Each sample was duplicated in two adjacent wells to allow the reproducibility of the arraying and hybridization process to be tested. Samples from the microtiter plate were printed onto glass microscope slides in an area measuring 3.5 mm by 5.5 mm with the use of a high-speed arraying machine (3). The arrays were processed by chemical and heat treatment to attach the DNA sequences to the glass surface and denature them (3). Three arrays, printed in a single lot, were used for the experiments here. A single microtiter plate of PCR products provides sufficient material to print at least 500 arrays.

Fluorescent probes were prepared from total *Arabidopsis* mRNA (4) by a single round of reverse transcription (5). The *Arabidopsis* mRNA was supplemented with human acetylcholine receptor (AChR) mRNA at a dilution of 1:10,000 (w/w) before cDNA synthesis, to provide an internal standard for calibration (5). The resulting fluorescently labeled cDNA mixture was hybridized to an array at high stringency (6) and scanned

M. Schena and R. W. Davis, Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305, USA.

D. Shalon and P. O. Brown, Department of Biochemistry and Howard Hughes Medical Institute, Beckman Center, Stanford University Medical Center, Stanford, CA 94305, USA.

\*These authors contributed equally to this work.

†Present address: Syntex, Palo Alto, CA 94303, USA.

‡To whom correspondence should be addressed. E-mail: pbrown@crgm.stanford.edu

24. J. Chant and I. Herskowitz, *Cell* 65, 1203 (1991).
25. B. W. Matthews, *Acc. Chem. Res.* 21, 333 (1988).
26. K. Kuchler, H. G. Dohman, J. Thorne, *J. Cell Biol.* 120, 1203 (1993); R. Kolling and C. P. Hollenback, *EMBO J.* 13, 3261 (1994); C. Berkower, D. Loeyza, S. Michalek, *Mol. Biol. Cell* 5, 1185 (1994).
27. A. Bender and J. R. Pringle, *Proc. Natl. Acad. Sci. USA* 86, 9976 (1989); J. Chant, K. Corrado, J. R. Pringle, I. Herskowitz, *Cell* 65, 1213 (1991); S. Powers, E. Gonzales, T. Christensen, J. Cubert, D. Broek, *ibid.*, p. 1225; H. O. Park, J. Chant, I. Herskowitz, *Nature* 365, 269 (1993); J. Chant, *Trends Genet.* 10, 328 (1994); ——— and J. R. Pringle, *J. Cell Biol.* 129, 751 (1995); J. Chant, M. Mischke, E. Mitchell, I. Herskowitz, J. R. Pringle, *ibid.*, p. 767.
28. G. F. Sprague Jr., *Methods. Enzymol.* 194, 77 (1991).
29. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
30. A W303 1A derivative, SY2625 (MATa *ura3-1 leu2-3, 112 trp1-1 ade2-1 can1-100 ss1Δ ma2Δ::URA3-1::his3Δ::URA3-HIS3*), was the parent strain for the mutant search. SY2625 derivatives for the mating assays, secreted pheromone assays, and the pulse-chase experiments included the following strains: Y49 (*ste22-1*), Y115 (*ma1Δ::LEU2*), Y142 (*ad1::URA3*), Y173 (*ad1Δ::LEU2*), Y220 (*ad1::URA3 ste23Δ::URA3*), Y221 (*ste23Δ::URA3*), Y231 (*ad1Δ::LEU2 ste23Δ::LEU2*), and Y233 (*ste23Δ::LEU2*). MATa derivatives of SY2625 included the following strains: Y189 (SY2625 made MATa), Y278 (*ste22-1*), Y195 (*ma1Δ::LEU2*), Y196 (*ad1Δ::LEU2*), and Y187 (*ad1::URA3*). The EG123 (MATa *leu2 ura3 trp1 can1 his4*) genetic background was used to create a set of strains for analysis of bud site selection. EG123 derivatives included the following strains: Y175 (*ad1Δ::LEU2*), Y223 (*ad1::URA3*), Y234 (*ste23Δ::LEU2*), and Y272 (*ad1Δ::LEU2 ste23Δ::LEU2*). MATa derivatives of EG123 included the following strains: Y214 (EG123 made MATa) and Y283 (*ad1Δ::LEU2*). All strains were generated by means of standard genetic or molecular methods involving the appropriate constructs (23). In particular, the *ad1 ste23* double mutant strains were created by crossing of the appropriate MATa *ste23* and MATa *ad1* mutants, followed by sporulation of the resultant diploid and isolation of the double mutant from nonparental di-type tetrads. Gene disruptions were confirmed with either PCR or Southern (DNA) analysis.
31. p129 is a YEp352 [J. E. Hill, A. M. Myers, T. J. Koerner, A. Tzagoloff, *Yeast* 2, 163 (1986)] plasmid containing a 5.5-kb Sal I fragment of pAX1. p151 was derived from p129 by insertion of a linker at the Bgl II site within AX1, which led to an in-frame insertion of the hemagglutinin (HA) epitope (DQYFYDPOYA) (29) between amino acids 854 and 855 of the AX1 prod-

with a laser (3). A high-sensitivity scan gave signals that saturated the detector at nearly all of the *Arabidopsis* target sites (Fig. 1A). Calibration relative to the AChR mRNA standard (Fig. 1A) established a sensitivity limit of  $\sim 1:50,000$ . No detectable hybridization was observed to either the rat glucocorticoid receptor (Fig. 1A) or the yeast TRP4 (Fig. 1A) targets even at the highest scanning sensitivity. A moderate-sensitivity scan

of the same array allowed linear detection of the more abundant transcripts (Fig. 1B). Quantitation of both scans revealed a range of expression levels spanning three orders of magnitude for the 45 genes tested (Table 2). RNA blots (7) for several genes (Fig. 2) corroborated the expression levels measured with the microarray to within a factor of 5 (Table 2).

Differential gene expression was investi-

gated with a simultaneous, two-color hybridization scheme, which served to minimize experimental variation inherent in the comparison of independent hybridizations. Fluorescent probes were prepared from two mRNA sources with the use of reverse transcriptase in the presence of fluorescein- and lissamine-labeled nucleotide analogs, respectively (5). The two probes were then mixed together in equal proportions, hybridized to a single array, and scanned separately for fluorescein and lissamine emission after independent excitation of the two fluorophores (3).

To test whether overexpression of a single gene could be detected in a pool of total *Arabidopsis* mRNA, we used a microarray to analyze a transgenic line overexpressing the single transcription factor HAT4 (8). Fluorescent probes representing mRNA from wild-type and HAT4-transgenic plants were labeled with fluorescein and lissamine, respectively; the two probes were then mixed and hybridized to a single array. An intense hybridization signal was observed at the position of the HAT4 cDNA in the lissamine-specific scan (Fig. 1D), but not in the fluorescein-specific scan of the same array (Fig. 1C). Calibration with AChR mRNA added to the fluorescein and lissamine cDNA synthesis reactions at dilutions of 1:10,000 (Fig. 1C) and 1:100 (Fig. 1D), respectively, revealed a 50-fold elevation of HAT4 mRNA in the transgenic line relative to its abundance in wild-type plants (Table 2). This magnitude of HAT4 overexpression matched that inferred from the Northern (RNA) analysis within a factor of 2 (Fig. 2 and Table 2). Expression of all the other genes monitored on the array differed by less than a factor of 5 between HAT4-transgenic and wild-type plants (Fig. 1, C

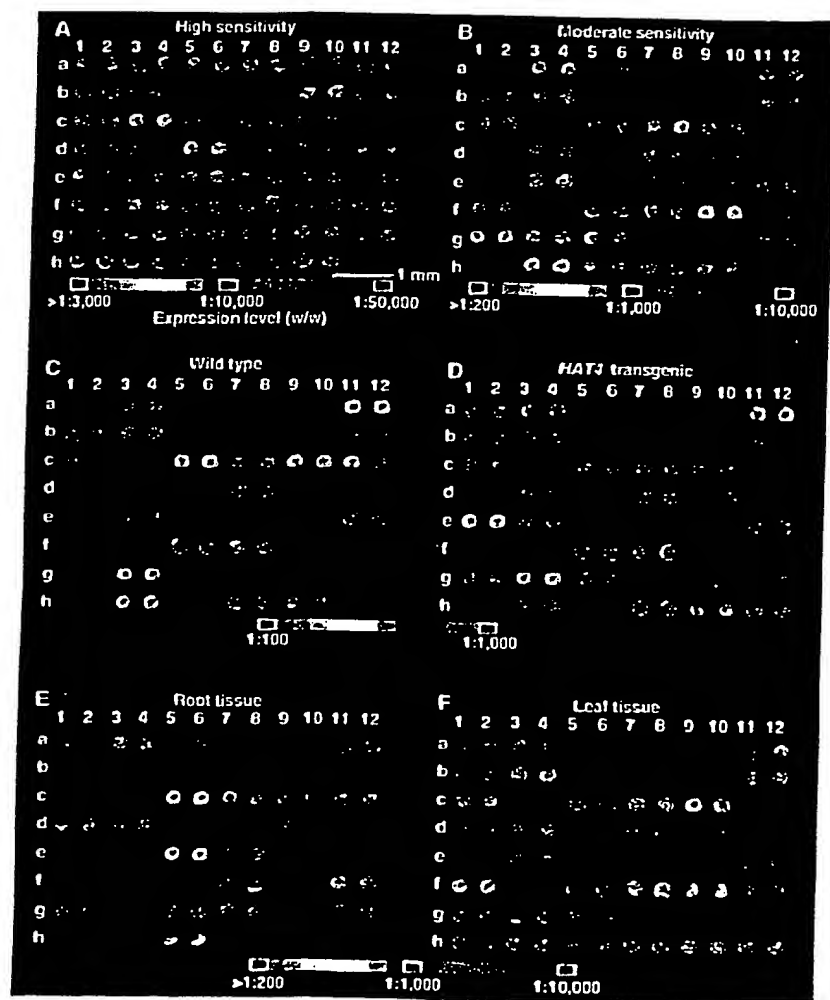


Fig. 1. Gene expression monitored with the use of cDNA microarrays. Fluorescent scans represented in pseudocolor correspond to hybridization intensities. Color bars were calibrated from the signal obtained with the use of known concentrations of human AChR mRNA in independent experiments. Numbers and letters on the axes mark the position of each cDNA. (A) High-sensitivity fluorescein scan after hybridization with fluorescein-labeled cDNA derived from wild-type plants. (B) Same array as in (A) but scanned at moderate sensitivity. (C and D) A single array was probed with a 1:1 mixture of fluorescein-labeled cDNA from wild-type plants and lissamine-labeled cDNA from HAT4-transgenic plants. The single array was then scanned successively to detect the fluorescein fluorescence corresponding to mRNA from wild-type plants (C) and the lissamine fluorescence corresponding to mRNA from HAT4-transgenic plants (D). (E and F) A single array was probed with a 1:1 mixture of fluorescein-labeled cDNA from root tissue and lissamine-labeled cDNA from leaf tissue. The single array was then scanned successively to detect the fluorescein fluorescence corresponding to mRNAs expressed in roots (E) and the lissamine fluorescence corresponding to mRNAs expressed in leaves (F).

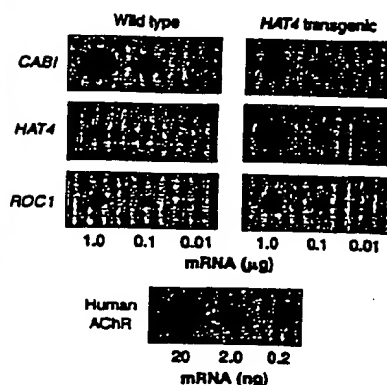


Fig. 2. Gene expression monitored with RNA (Northern) blot analysis. Designated amounts of mRNA from wild-type and HAT4-transgenic plants were spotted onto nylon membranes and probed with the cDNAs indicated. Purified human AChR mRNA was used for calibration.

and D, and Table 2). Hybridization of fluorescein-labeled glucocorticoid receptor cDNA (Fig. 1C) and lissamine-labeled TRP4 cDNA (Fig. 1D) verified the presence of the negative control targets and the lack of optical cross talk between the two fluorophores.

To explore a more complex alteration in expression patterns, we performed a second two-color hybridization experiment with fluorescein- and lissamine-labeled probes prepared from root and leaf mRNA, respectively. The scanning sensitivities for the two fluorophores were normalized by matching the signals resulting from AChR

mRNA, which was added to both cDNA synthesis reactions at a dilution of 1:1000 (Fig. 1, E and F). A comparison of the scans revealed widespread differences in gene expression between root and leaf tissue (Fig. 1, E and F). The mRNA from the light-regulated CAB1 gene was ~500-fold more abundant in leaf (Fig. 1F) than in root tissue (Fig. 1E). The expression of 26 other genes differed between root and leaf tissue by more than a factor of 5 (Fig. 1, E and F).

The HAT4-transgenic line we examined has elongated hypocotyls, early flowering, poor germination, and altered pigmentation (8). Although changes in expression were

observed for HAT4, large changes in expression were not observed for any of the other 44 genes we examined. This was somewhat surprising, particularly because comparative analysis of leaf and root tissue identified 27 differentially expressed genes. Analysis of an expanded set of genes may be required to identify genes whose expression changes upon HAT4 overexpression; alternatively, a comparison of mRNA populations from specific tissues of wild-type and HAT4-transgenic plants may allow identification of downstream genes.

At the current density of robotic printing, it is feasible to scale up the fabrication process to produce arrays containing 20,000 cDNA targets. At this density, a single array would be sufficient to provide gene-specific targets encompassing nearly the entire repertoire of expressed genes in the *Arabidopsis* genome (2). The availability of 20,274 ESTs from *Arabidopsis* (1, 9) would provide a rich source of templates for such studies.

The estimated 100,000 genes in the human genome (10) exceeds the number of *Arabidopsis* genes by a factor of 5 (2). This modest increase in complexity suggests that similar cDNA microarrays, prepared from the rapidly growing repertoire of human ESTs (1), could be used to determine the expression patterns of tens of thousands of human genes in diverse cell types. Coupling an amplification strategy to the reverse transcription reaction (11) could make it feasible to monitor expression even in minute tissue samples. A wide variety of acute and chronic physiological and pathological conditions might lead to characteristic changes in the patterns of gene expression in peripheral blood cells or other easily sampled tissues. In concert with cDNA microarrays for monitoring complex expression patterns, these tissues might therefore serve as sensitive *in vivo* sensors for clinical diagnosis. Microarrays of cDNAs could thus provide a useful link between human gene sequences and clinical medicine.

Table 2. Gene expression monitoring by microarray and RNA blot analyses; tg, HAT4-transgenic. See Table 1 for additional gene information. Expression levels (w/w) were calibrated with the use of known amounts of human AChR mRNA. Values for the microarray were determined from microarray scans (Fig. 1); values for the RNA blot were determined from RNA blots (Fig. 2).

Gene	Expression level (w/w)	
	Microarray	RNA blot
CAB1	1:48	1:83
CAB1 (tg)	1:120	1:150
HAT4	1:8300	1:6300
HAT4 (tg)	1:150	1:210
ROC1	1:1200	1:1800
ROC1 (tg)	1:260	1:1300

Table 1. Sequences contained on the cDNA microarray. Shown is the position, the known or putative function, and the accession number of each cDNA in the microarray (Fig. 1). All but three of the ESTs used in this study matched a sequence in the database. NADH, reduced form of nicotinamide adenine dinucleotide; ATPase, adenosine triphosphatase; GTP, guanosine triphosphate.

Position	cDNA	Function	Accession number
a1, 2	AChR	Human AChR	-
a3, 4	EST3	Actin	H36236
a5, 6	EST6	NADH dehydrogenase	Z27010
a7, 8	AAC1	Actin 1	M20016
a9, 10	EST12	Unknown	U365941
a11, 12	EST13	Actin	T45783
b1, 2	CAB1	Chlorophyll a/b binding	M85150
b3, 4	EST17	Phosphoglycerate kinase	T44490
b5, 6	GA4	Gibberellic acid biosynthesis	L37126
b7, 8	EST19	Unknown	U365951
b9, 10	GBF-1	G-box binding factor 1	X63894
b11, 12	EST23	Elongation factor	X52256
c1, 2	EST29	Aldolase	T04477
c3, 4	GBF-2	G-box binding factor 2	X63895
c5, 6	EST34	Chloroplast protease	R87034
c7, 8	EST35	Unknown	T14152
c9, 10	EST41	Catalase	T22720
c11, 12	rGR	Rat glucocorticoid receptor	M14053
d1, 2	EST42	Unknown	U365961
d3, 4	EST45	ATPase	J04185
d5, 6	HAT1	Homeobox-leucine zipper 1	U09332
d7, 8	EST46	Light harvesting complex	T04063
d9, 10	EST49	Unknown	T78267
d11, 12	HAT2	Homeobox-leucine zipper 2	U09335
e1, 2	HAT4	Homeobox-leucine zipper 4	M90394
e3, 4	EST50	Phosphoributokinase	T04344
e5, 6	HAT5	Homeobox-leucine zipper 5	M90416
e7, 8	EST51	Unknown	Z33675
e9, 10	HAT22	Homeobox-leucine zipper 22	U09336
e11, 12	EST52	Oxygen evolving	T21749
f1, 2	EST59	Unknown	Z34607
f3, 4	KNAT1	Knotted-like homeobox 1	U14174
f5, 6	EST60	RuBisCO small subunit	X14564
f7, 8	EST69	Translation elongation factor	T42799
f9, 10	PPH1	Protein phosphatase 1	U34803
f11, 12	EST70	Unknown	T44621
g1, 2	EST75	Chloroplast protease	T43698
g3, 4	EST78	Unknown	R65481
g5, 6	ROC1	Cyclophilin	L14844
g7, 8	EST82	GTP binding	X59152
g9, 10	EST83	Unknown	Z33795
g11, 12	EST84	Unknown	T45278
h1, 2	EST91	Unknown	T13832
h3, 4	EST96	Unknown	R64816
h5, 6	SAR1	Synaptobrevin	M90418
h7, 8	EST100	Light harvesting complex	Z18205
h9, 10	EST103	Light harvesting complex	X03909
h11, 12	TRP4	Yeast tryptophan biosynthesis	X04273

\*Proprietary sequence of Stratagene (La Jolla, California).

†No match in the database; novel EST.

## REFERENCES AND NOTES

1. The current EST database (dbEST release 091495) from the National Center for Biotechnology Information (Bethesda, MD) contains a total of 322,225 entries, including 255,645 from the human genome and 21,044 from Arabidopsis. Access is available via the World Wide Web (<http://www.ncbi.nlm.nih.gov>).
2. E. M. Meyerowitz and R. E. Pruitt, *Science* 229, 1214 (1985); R. E. Pruitt and E. M. Meyerowitz, *J. Mol. Biol.* 187, 169 (1986); L. Hwang et al., *Plant J.* 1, 367 (1991); P. Jarvis et al., *Plant Mol. Biol.* 24, 685 (1994); L. Le Guen et al., *Mol. Gen. Genet.* 245, 390 (1994).
3. D. Shelton, thesis, Stanford University (1995); and P. O. Brown, in preparation. Microarrays were fabricated on poly-L-lysine-coated microscope slides (Sigma) with a custom-built arraying machine fitted with one printing tip. The tip loaded 1  $\mu$ l of PCR product (0.5 mg/ml) from 96-well microtiter plates and deposited  $\sim$ 0.005  $\mu$ l per slide on 40 slides at a spacing of 500  $\mu$ m. The printed slides were rehydrated for 2 hours in a humid chamber, snap-dried at 100°C for 1 min, rinsed in 0.1% SDS, and treated with 0.05% succinic anhydride prepared in buffer consisting of 50% 1-methyl-2-pyrrolidone and 50% boric acid. The cDNA on the slides was denatured in distilled water for 2 min at 90°C immediately before use. Microarrays were scanned with a laser fluorescent scanner that contained a computer-controlled XY stage and a microscope objective. A mixed gas, multiline laser allowed sequential excitation of the two fluorophores. Emitted light was split according to wavelength and detected with two photomultiplier tubes. Signals were read into a PC with the use of a 12-bit analog-to-digital board. Additional details of microarray fabrication and use may be obtained by means of e-mail ([pbrown@cormn.stanford.edu](mailto:pbrown@cormn.stanford.edu)).
4. F. M. Ausubel et al., Eds., *Current Protocols in Molecular Biology* (Greene & Wiley Interscience, New York, 1994), pp. 4.3.1–4.3.4.
5. Polyadenylated [poly(A)<sup>+</sup>] mRNA was prepared from total RNA with the use of Oligotex-dT resin (Qiagen). Reverse transcription (RT) reactions were carried out with a Stratascript RT-PCR kit (Stratagene) modified as follows: 50- $\mu$ l reactions contained 0.1  $\mu$ g/ $\mu$ l of Arabidopsis mRNA, 0.1 ng/ $\mu$ l of human AChR mRNA, 0.05  $\mu$ g/ $\mu$ l of oligo(dT) (21-mer), 1 $\times$  first strand buffer, 0.03 U/ $\mu$ l of ribonuclease block, 500  $\mu$ M deoxyadenosine triphosphate (dATP), 500  $\mu$ M deoxycytidine triphosphate, 500  $\mu$ M dTTP, 40  $\mu$ M deoxycytosine triphosphate (dCTP), 40  $\mu$ M fluorescein-12-dCTP (or fluorescein-5-dCTP), and 0.03 U/ $\mu$ l of Stratascript reverse transcriptase. Reactions were incubated for 60 min at 37°C, precipitated with ethanol, and resuspended in 10  $\mu$ l of TE (10 mM Tris-HCl and 1 mM EDTA, pH 8.0). Samples were then heated for 3 min at 94°C and chilled on ice. The RNA was degraded by adding 0.25  $\mu$ l of 10 N NaOH followed by a 10-min incubation at 37°C. The samples were neutralized by addition of 2.5  $\mu$ l of 1 M Tris-Cl (pH 8.0) and 0.25  $\mu$ l of 10 N HCl and precipitated with ethanol. Pellets were washed with 70% ethanol, dried to completion in a speedvac, resuspended in 10  $\mu$ l of H<sub>2</sub>O, and reduced to 3.0  $\mu$ l in a speedvac. Fluorescent nucleotide analogs were obtained from New England Nuclear (DuPont).
6. Hybridization reactions contained 1.0  $\mu$ l of fluorescent cDNA synthesis product (5) and 1.0  $\mu$ l of hybridization buffer (10 $\times$  saline sodium citrate (SSC) and 0.2% SDS). The 2.0- $\mu$ l probe mixtures were aliquoted onto the microarray surface and covered with cover slips (12 mm round). Arrays were transferred to a hybridization chamber (3) and incubated for 18 hours at 65°C. Arrays were washed for 5 min at room temperature (25°C) in low-stringency wash buffer (1 $\times$  SSC and 0.1% SDS), then for 10 min at room temperature in high-stringency wash buffer (0.1 $\times$  SSC and 0.1% SDS). Arrays were scanned in 0.1 $\times$  SSC with the use of a fluorescence laser-scanning device (3).
7. Samples of poly(A)<sup>+</sup> mRNA (4, 5) were spotted onto nylon membranes (Hytran) and crosslinked with ultraviolet light with the use of a Stratilinker 1800 (Stratagene). Probes were prepared by random priming with the use of a Prime-It II kit (Stratagene) in the presence of [<sup>32</sup>P]dATP. Hybridizations were carried out according to the instructions of the manufacturer. Quantitation was performed on a PhosphorImager (Molecular Dynamics).
8. M. Schena and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* 89, 3894 (1992); M. Schena, A. M. Lloyd, R. W. Davis, *Genes Dev.* 7, 367 (1993); M. Schena and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* 91, 8393 (1994).
9. H. Hofte et al., *Plant J.* 4, 1051 (1993); T. Newman et al., *Plant Physiol.* 106, 1241 (1994).
10. N. E. Morton, *Proc. Natl. Acad. Sci. U.S.A.* 88, 7474 (1991); E. D. Green and R. H. Waterston, *J. Am. Med. Assoc.* 266, 1966 (1991); C. Belletto-Chentelot, *Cell* 70, 1059 (1992); D. R. Cox et al., *Science* 265, 2031 (1994).
11. E. S. Kawasaki et al., *Proc. Natl. Acad. Sci. U.S.A.* 85, 5688 (1988).
12. The laser fluorescent scanner was designed and fabricated in collaboration with S. Smith of Stanford University. Scanner and analysis software was developed by R. X. Xia. The succinic anhydride reaction was suggested by J. Muligan and J. Van Ness of Darwin Molecular Corporation. Thanks to S. Theologos, C. Somerville, K. Yamamoto, and members of the laboratories of R.W.D. and P.O.B. for critical comments. Supported by the Howard Hughes Medical Institute and by grants from NIH (R21HG00450) (P.O.B.) and R01AG00198 (R.W.D.) and from NSF (MCB9108011) (R.W.D.) and by an NSF graduate fellowship (D.S.). P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

11 August 1995; accepted 22 September 1995

## Gene Therapy in Peripheral Blood Lymphocytes and Bone Marrow for ADA<sup>-</sup> Immunodeficient Patients

Claudio Bordignon,\* Luigi D. Notarangelo, Nadia Nobili, Giuliana Ferrari, Giulia Casorati, Paola Panina, Evelina Mazzolari, Daniela Maggioni, Claudia Rossi, Paolo Servida, Alberto G. Ugazio, Fulvio Mavilio

Adenosine deaminase (ADA) deficiency results in severe combined immunodeficiency, the first genetic disorder treated by gene therapy. Two different retroviral vectors were used to transfer ex vivo the human ADA minigene into bone marrow cells and peripheral blood lymphocytes from two patients undergoing exogenous enzyme replacement therapy. After 2 years of treatment, long-term survival of T and B lymphocytes, marrow cells, and granulocytes expressing the transferred ADA gene was demonstrated and resulted in normalization of the immune repertoire and restoration of cellular and humoral immunity. After discontinuation of treatment, T lymphocytes, derived from transduced peripheral blood lymphocytes, were progressively replaced by marrow-derived T cells in both patients. These results indicate successful gene transfer into long-lasting progenitor cells, producing a functional multilineage progeny.

Severe combined immunodeficiency associated with inherited deficiency of ADA (1) is usually fatal unless affected children are kept in protective isolation or the immune system is reconstituted by bone marrow transplantation from a human leukocyte antigen (HLA)-identical sibling donor (2). This is the therapy of choice, although it is available only for a minority of patients. In recent years, other forms of therapy have been developed, including transplants from haploidentical donors (3, 4), exogenous enzyme replacement (5), and somatic-cell gene therapy (6–9).

We previously reported a preclinical model in which ADA gene transfer and expression

successfully restored immune functions in human ADA-deficient (ADA<sup>-</sup>) peripheral blood lymphocytes (PBLs) in immunodeficient mice in vivo (10, 11). On the basis of these preclinical results, the clinical application of gene therapy for the treatment of ADA<sup>-</sup> SCID (severe combined immunodeficiency disease) patients who previously failed exogenous enzyme replacement therapy was approved by our Institutional Ethical Committees and by the Italian National Committee for Bioethics (12). In addition to evaluating the safety and efficacy of the gene therapy procedure, the aim of the study was to define the relative role of PBLs and hematopoietic stem cells in the long-term reconstitution of immune functions after retroviral vector-mediated ADA gene transfer. For this purpose, two structurally identical vectors expressing the human ADA complementary DNA (cDNA), distinguishable by the presence of alternative restriction sites in a nonfunctional region of the viral long-terminal repeat (LTR), were used to transduce PBLs and bone marrow (BM) cells independently. This procedure allowed identification of the origin of

C. Bordignon, N. Nobili, G. Ferrari, D. Maggioni, C. Rossi, P. Servida, F. Mavilio, Telethon Gene Therapy Program for Genetic Diseases, DIBT, Istituto Scientifico H. S. Raffaele, Milan, Italy.  
L. D. Notarangelo, E. Mazzolari, A. G. Ugazio, Department of Pediatrics, University of Brescia Medical School, Brescia, Italy.  
G. Casorati, Unità di Immunochimica, DIBT, Istituto Scientifico H. S. Raffaele, Milan, Italy.  
P. Panina, Roche Milano Ricerche, Milan, Italy.

\*To whom correspondence should be addressed.

## Discovery and analysis of inflammatory disease-related genes using cDNA microarrays

(Inflammation/human genome analysis/gene discovery)

RENU A. HELLER<sup>†</sup>, MARK SCHENA<sup>\*</sup>, ANDREW CHAI<sup>\*</sup>, DARI SHALON<sup>‡</sup>, TOD BEDILION<sup>‡</sup>, JAMES GILMORE<sup>‡</sup>, DAVID E. WOOLLEY<sup>§</sup>, AND RONALD W. DAVIS<sup>\*</sup>

<sup>\*</sup>Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305; <sup>‡</sup>Synteni, Palo Alto, CA 94306; and <sup>§</sup>Department of Medicine, Manchester Royal Infirmary, Manchester, United Kingdom

Contributed by Ronald W. Davis, December 27, 1996

**ABSTRACT** cDNA microarray technology is used to profile complex diseases and discover novel disease-related genes. In inflammatory disease such as rheumatoid arthritis, expression patterns of diverse cell types contribute to the pathology. We have monitored gene expression in this disease state with a microarray of selected human genes of probable significance in inflammation as well as with genes expressed in peripheral human blood cells. Messenger RNA from cultured macrophages, chondrocyte cell lines, primary chondrocytes, and synovial cells provided expression profiles for the selected cytokines, chemokines, DNA binding proteins, and matrix-degrading metalloproteinases. Comparisons between tissue samples of rheumatoid arthritis and inflammatory bowel disease verified the involvement of many genes and revealed novel participation of the cytokine interleukin 3, chemokine Gro $\alpha$  and the metalloproteinase matrix metallo-elastase in both diseases. From the peripheral blood library, tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase genes were identified as expressed differentially in rheumatoid arthritis compared with inflammatory bowel disease. These results successfully demonstrate the use of the cDNA microarray system as a general approach for dissecting human diseases.

The recently described cDNA microarray or DNA-chip technology allows expression monitoring of hundreds and thousands of genes simultaneously and provides a format for identifying genes as well as changes in their activity (1, 2). Using this technology, two-color fluorescence patterns of differential gene expression in the root versus the shoot tissue of *Arabidopsis* were obtained in a specific array of 48 genes (1). In another study using a 1000 gene array from a human peripheral blood library, novel genes expressed by T cells were identified upon heat shock and protein kinase C activation (3).

The technology uses cDNA sequences or cDNA inserts of a library for PCR amplification that are arrayed on a glass slide with high speed robotics at a density of 1000 cDNA sequences per cm<sup>2</sup>. These microarrays serve as gene targets for hybridization to cDNA probes prepared from RNA samples of cells or tissues. A two-color fluorescence labeling technique is used in the preparation of the cDNA probes such that a simultaneous hybridization but separate detection of signals provides the comparative analysis and the relative abundance of specific genes expressed (1, 2). Microarrays can be constructed from specific cDNA clones of interest, a cDNA library, or a select number of open reading frames from a genome sequencing database to allow a large-scale functional analysis of expressed sequences.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA  
0027-8424/97/942150-06\$2.00/0  
PNAS is available online at <http://www.pnas.org>.

Because of the wide spectrum of genes and endogenous mediators involved, the microarray technology is well suited for analyzing chronic diseases. In rheumatoid arthritis (RA), inflammation of the joint is caused by the gene products of many different cell types present in the synovium and cartilage tissues plus those infiltrating from the circulating blood. The autoimmune and inflammatory nature of the disease is a cumulative result of genetic susceptibility factors and multiple responses, paracrine and autocrine in nature, from macrophages, T cells, plasma cells, neutrophils, synovial fibroblasts, chondrocytes, etc. Growth factors, inflammatory cytokines (4), and the chemokines (5) are the important mediators of this inflammatory process. The ensuing destruction of the cartilage and bone by the invading synovial tissue includes the actions of prostaglandins and leukotrienes (6), and the matrix-degrading metalloproteinases (MMPs). The MMPs are an important class of Zn-dependent metallo-endoproteases that can collectively degrade the proteoglycan and collagen components of the connective tissue matrix (7).

This paper presents a study in which the involvement of select classes of molecules in RA was examined. Also investigated were 1000 human genes randomly selected from a peripheral human blood cell library. Their differential and quantitative expression analysis in cells of the joint tissue, in diseased RA tissue and in inflammatory bowel disease (IBD) tissues was conducted to demonstrate the utility of the microarray method to analyze complex diseases by their pattern of gene expression. Such a survey provides insight not only into the underlying cause of the pathology, but also provides the opportunity to selectively target genes for disease intervention by appropriate drug development and gene therapies.

### METHODS

**Microarray Design, Development, and Preparation.** Two approaches for the fabrication of cDNA microarrays were used in this study. In the first approach, known human genes of probable significance in RA were identified. Regions of the clones, preferably 1 kb in length, were selected by their proximity to the 3' end of the cDNA and for areas of least identity to related and repetitive sequences. Primers were synthesized to amplify the target regions by standard PCR protocols (3). Products were

Abbreviations: RA, rheumatoid arthritis; MMP, matrix-degrading metalloproteinase; IBD, inflammatory bowel disease; LPS, lipopolysaccharide; PMA, phorbol 12-myristate 13-acetate; TNF- $\alpha$ , tumor necrosis factor  $\alpha$ ; IL, interleukin; TGF- $\beta$ , transforming growth factor  $\beta$ ; G-CSF, granulocyte colony-stimulating factor; MIP, macrophage inflammatory protein; MIF, migration inhibitory factor; HME, human matrix metallo-elastase; RANTES, regulated upon activation, normal T cell expressed and secreted; Gel, gelatinase; VCAM, vascular cell adhesion molecule; ICE, IL-1 converting enzyme; PUMP, putative metalloproteinase; MnSOD, manganese superoxide dismutase; TIMP, tissue inhibitor of metalloproteinase; MCP, macrophage chemotactic protein.

<sup>†</sup>To whom reprint requests should be sent at the present address: Roche Bioscience, S3-1, 3401 Hillview Avenue, Palo Alto, CA 94304.

verified by gel electrophoresis and purified with Qiaquick 96-well purification kit (Qiagen, Chatsworth, CA), lyophilized (Savant), and resuspended in 5  $\mu$ l of 3 $\times$  standard saline citrate (SSC) buffer for arraying. In the second approach, the microarray containing the 1056 human genes from the peripheral blood lymphocyte library was prepared as described (3).

**Tissue Specimens.** Rheumatoid synovial tissue was obtained from patients with late stage classic RA undergoing remedial synovectomy or arthroplasty of the knee. Synovial tissue was separated from any associated connective tissue or fat. One gram of each synovial specimen was subjected to RNA extraction within 40 min of surgical excision, or explants were cultured in serum-free medium to examine any changes under *in vitro* conditions. For IBD, specimens of macroscopically inflamed lower intestinal mucosa were obtained from patients with Crohn disease undergoing remedial surgery. The hypertrophied mucosal tissue was separated from underlying connective tissue and extracted for RNA.

**Cultured Cells.** The Mono Mac-6 (MM6) monocytic cells (8) were grown in RPMI medium. Human chondrosarcoma SW1353 cells, primary human chondrocytes, and synoviocytes (9, 10) were cultured in DMEM; all culture media were supplemented with 10% fetal bovine serum, 100  $\mu$ g/ml streptomycin, and 500 units/ml penicillin. Treatment of cells with lipopolysaccharide (LPS) endotoxin at 30 ng/ml, phorbol 12-myristate 13-acetate (PMA) at 50 ng/ml, tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ) at 50 ng/ml, interleukin (IL)-1 $\beta$  at 30 ng/ml, or transforming growth factor- $\beta$  (TGF- $\beta$ ) at 100 ng/ml is described in the figure legends.

**Fluorescent Probe, Hybridization, and Scanning.** Isolation of mRNA, probe preparation, and quantitation with *Arabidopsis* control mRNAs was essentially as described (3) except for the following minor modification. Following the reverse transcriptase step, the appropriate Cy3- and Cy5-labeled samples were pooled; mRNA degraded by heating the sample to 65°C for 10 min with the addition of 5  $\mu$ l of 0.5M NaOH plus 0.5 ml of 10 mM EDTA. The pooled cDNA was purified from unincorporated nucleotides by gel filtration in Centri-spin columns (Princeton Separations, Adelphia, NJ). Samples were lyophilized and dissolved in 6  $\mu$ l of hybridization buffer (5 $\times$  SSC plus 0.2% SDS). Hybridizations, washes, scanning, quantitation procedures, and pseudocolor representations of fluorescent images have been described (3). Scans for the two fluorescent probes were normalized either to the fluorescence intensity of *Arabidopsis* mRNAs spiked into the labeling reactions (see Figs. 2–4) or to the signal intensity of  $\beta$ -actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH; see Fig. 5).

## RESULTS

**Ninety-Six-Genes Microarray Design.** The actions of cytokines, growth factors, chemokines, transcription factors, MMPs, prostaglandins, and leukotrienes are well recognized in inflammatory disease, particularly RA (11–14). Fig. 1 displays the selected genes for this study and also includes control cDNAs of housekeeping genes such as  $\beta$ -actin and GAPDH and genes from *Arabidopsis* for signal normalization and quantitation (row A, columns 1–12).

**Defining Microarray Assay Conditions.** Different lengths and concentrations of target DNA were tested by arraying PCR-

	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	BLANK	HAT1 HAT1	HAT1 HAT1	HAT4 HAT4	HAT4 HAT4	HAT22 HAT22	HAT22 HAT22	YES23 YES23	YES23 YES23	BACTIN $\beta$ -actin	G3PDH G3PDH
B	IL1A IL-1 $\alpha$	IL1B IL-1 $\beta$	IL1RA IL-1RA	IL2 IL-2	IL3 IL-3	IL4 IL-4	IL6 IL-6	IL6R IL-6R	IL7 IL-7	CFOS c-fos	CJUN c-jun	RFRA1 Rat Fra-1
C	IL8 IL-8	IL9 IL-9	IL10 IL-10	ICE ICE	IFNG IFN $\gamma$	GCSF G-CSF	MCSF M-CSF	GMCSF GM-CSF	TNFB.1 TNF $\beta$	CREL c-rel	NFKB50 NF $\kappa$ Bp50	NFKB65.1 NF $\kappa$ Bp65
D	TNFA.1 TNF $\alpha$	TNFA.2 TNF $\alpha$	TNFA.3 TNF $\alpha$	TNFA.4 TNF $\alpha$	TNFA.5 TNF $\alpha$	TNFR1.1 TNFR1	TNFR1.2 TNFR1	TNFR1.1 TNFR1	TNFR1.2 TNFR1	NFKB65.2 NF $\kappa$ Bp65	IKB I $\kappa$ B	CREB2 CREB2
E	STR1 Strom-1	STR2-3' Strom-2	STR3 Strom-3	COL1 Coll-1	COL1-3' Coll-1.3'	COL2.1 Coll-2	COL2.2 Coll-2	COL3 Coll-3	COX1 Cox-1	COX2 Cox-2	12LO 12-LO	15LO 15-LO
F	GELA.1 Gel-A	GELB Gel-B	HME Elastase	MTMMP MT-MMP	PUMP1 Matrilysin	TIMP1 TIMP-1	TIMP2 TIMP-2	TIMP3 TIMP-3	ICAM1 ICAM-1	VCAM VCAM	5LO.1 5-LO	CPLA2.2 cPLA2
G	EGF EGF	FGFA FGFacidic	FGFB FGFbasic	IGF1 IGF-1	IGF1 IGF-1	TGFA TGF $\alpha$	TGFB TGF $\beta$	PDGFB PDGF $\beta$	CALCTN Calcitonin	GH1 GH-1	GRO GRO1 $\alpha$	GCT GCT
H	MCP1.1 MCP-1	MCP1.1 MCP-1	MIP1A MIP-1 $\alpha$	MIP1B MIP-1 $\beta$	MIF MIF	RANTES RANTES	INOS INOS	LDLR LDLR	ALU.1 IL-10	ALU.2 TNFRp70	ALU.3 IL-10	POLYA LDLR

A. thaliana controls

Human controls

Cytokines and related genes

Transcription factors and related genes

MMP's and related genes

Chemokines

Growth factors and related genes

Other genes

FIG. 1. Ninety-six-element microarray design. The target element name and the corresponding gene are shown in the layout. Some genes have more than one target element to guarantee specificity of signal. For TNF the targets represent decreasing lengths of 1, 0.8, 0.6, 0.4, and 0.2 kb from left to right.



amplified products ranging from 0.2 to 1.2 kb at concentrations of 1  $\mu\text{g}/\mu\text{l}$  or less. No significant difference in the signal levels was observed within this range of target size and only with 0.2-kb length was a signal reduced upon an 8-fold dilution of the 1  $\mu\text{g}/\mu\text{l}$  sample (data not shown). In this study the average length of the targets was 1 kb, with a few exceptions in the range of  $\approx 300$  bp, arrayed at a concentration of 1  $\mu\text{g}/\mu\text{l}$ . Normally one PCR provided sufficient material to fabricate up to 1000 microarray targets.

In considering positional effects in the development of the targets for the microarrays, selection was biased toward the 3' proximal regions, because the signal was reduced if the target fragment was biased toward the 5' end (data not shown). This result was anticipated since the hybridizing probe is prepared by reverse transcription with oligo(dT)-primed mRNA and is richer in 3' proximal sequences. Cross-hybridizations of probes to targets of a gene family were analyzed with the matrix metal-

loproteinases as the example because they can show regions of sequence identities of greater than 70%. With collagenase-1 (Col-1) and collagenase-2 (Col-2) genes as targets with up to 70% sequence identity, and stromelysin-1 (Strom-1) and stromelysin-2 (Strom-2) genes with different degrees of identity, our results showed that a short region of overlap, even with 70–90% sequence identity, produced a low level of cross-hybridization. However, shorter regions of identity spread over the length of the target resulted in cross-hybridization (data not shown). For closely related genes, targets were designed by avoiding long stretches of homology. For members of a gene family two or more target regions were included to discriminate between specificity of signal versus cross-hybridization.

**Monitoring Differential Expression in Cultured Cell Lines.** In RA tissue, the monocyte/macrophage population plays a prominent role in phagocytic and immunomodulatory activities. Typ-

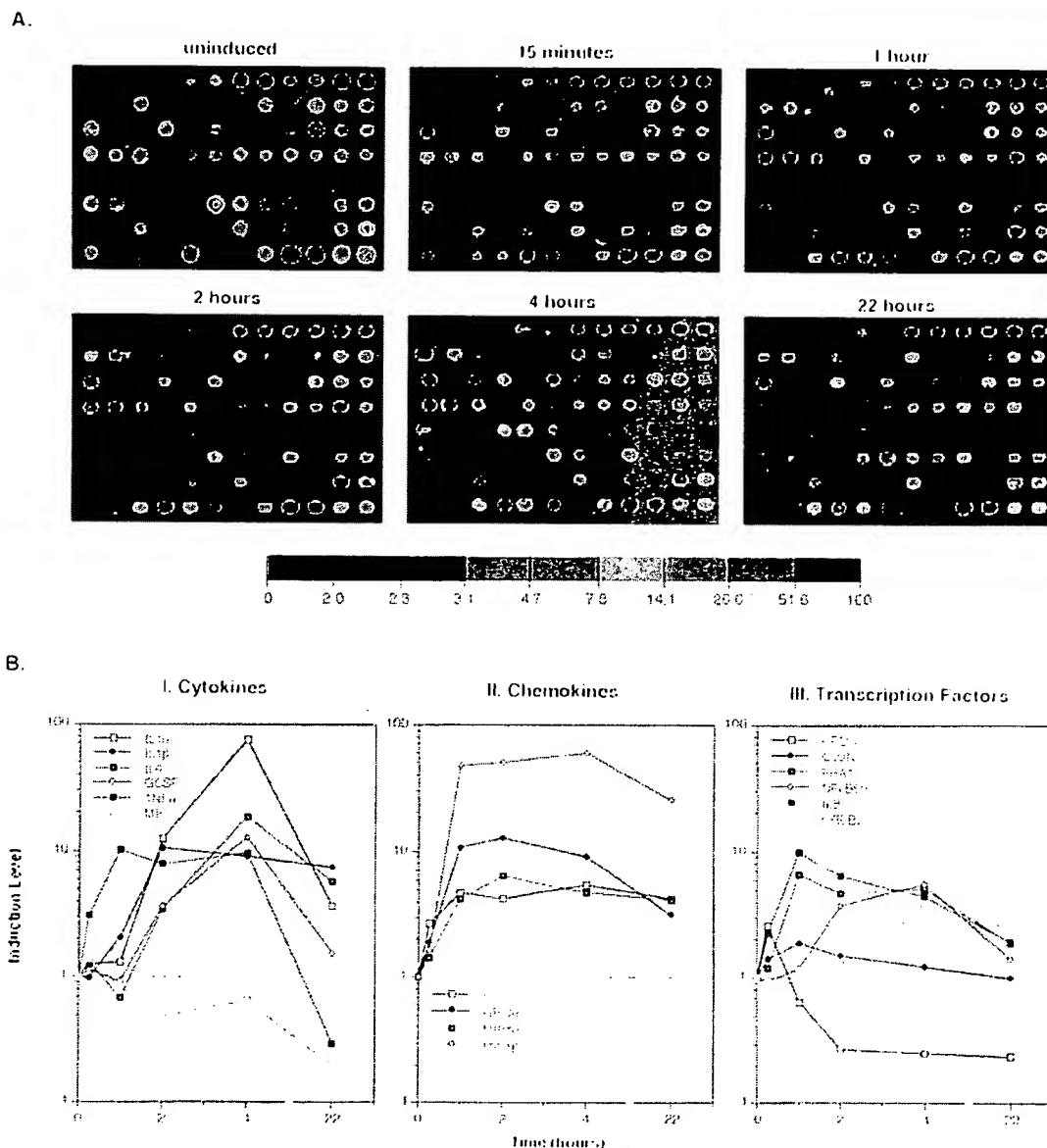


FIG. 2. Time course for LPS/PMA-induced MM6 cells. Array elements are described in Fig. 1. (A) Pseudocolor representations of fluorescent scans correspond to gene expression levels at each time point. The array is made up of 8 *Arabidopsis* control targets and 86 human cDNA targets, the majority of which are genes with known or suspected involvement in inflammation. The color bars provide a comparative calibration scale between arrays and are derived from the *Arabidopsis* mRNA samples that are introduced in equal amounts during probe preparation. Fluorescent probes were made by labeling mRNA from untreated MM6 cells or LPS and PMA treated cells. mRNA was isolated at indicated times after induction. (B I–III) The two-color samples were cohybridized, and microarray scans provided the data for the levels of select transcripts at different time points relative to abundance at time zero. The analysis was performed using normalized data collected from 8-bit images.



ically these cells, when triggered by an immunogen, produce the proinflammatory cytokines TNF and IL-1. We have used the monocyte cell line MM6 and monitored changes in gene expression upon activation with LPS endotoxin, a component of Gram-negative bacterial membranes, and PMA, which augments the action of LPS on TNF production (15). RNA was isolated at different times after induction and used for cDNA probe preparation. From this time course it was clear that TNF expression was induced within 15 min of treatment, reached maximum levels in 1 hr, remained high until 4 hr and subsequently declined (Fig. 2A). Many other cytokine genes were also transiently activated, such as IL-1 $\alpha$  and - $\beta$ , IL-6, and granulocyte colony-stimulating factor (G-CSF). Prominent chemokines activated were IL-8, macrophage inflammatory protein (MIP)-1 $\beta$ , more so than MIP-1 $\alpha$ , and Gro $\alpha$  or melanoma growth stimulatory factor. Migration inhibitory factor (MIF) expressed in the uninduced state declined in LPS-activated cells. Of the immediate early genes, the noticeable ones were *c-fos*, *fra-1*, *c-jun*, NF- $\kappa$ Bp50, and I $\kappa$ B, with *c-rel* expression observed even in the uninduced state (Fig. 2B). These expression patterns are consistent with reported patterns of activation of certain LPS- and PMA-induced genes (12). Demonstrated here is the unique ability of this system to allow parallel visualization of a large number of gene activities over a period of time.

SW1353 cells is a line derived from malignant tumors of the cartilage and behaves much like the chondrocytes upon stimulation with TNF and IL-1 in the expression of MMPs (9). In addition to confirming our earlier observations with Northern blots on Strom-1, Col-1, and Col-3 expression (9), gelatinase (Gel) A, putative metalloproteinase (PUMP)-1 membrane-

type matrix metalloproteinase, tissue inhibitors of matrix metalloproteinases or tissue inhibitor of metalloproteinase 1 (TIMP-1), -2, and -3 were also expressed by these cells together with the human matrix metallo-elastase (HME; Fig. 3A). HME induction was estimated to be  $\approx$ 50-fold and was greater than any of the other MMPs examined (Fig. 3B). This result was unexpected because HME is reportedly expressed only by alveolar macrophage and placental cells (16). Expression of the cytokines and chemokines, IL-6, IL-8, MIF, and MIP-1 $\beta$  was also noted. A variety of other genes, including certain transcription factors, were also up-regulated (Fig. 3), but the overall time-dependent expression of genes in the SW1353 cells was qualitatively distinct from the MM6 cells.

Quantitation of differential gene expression (Figs. 2B and 3B) was achieved with the simultaneous hybridization of Cy3-labeled cDNA from untreated cells and Cy5-labeled cDNA from treated samples. The estimated increases in expression from these microarrays for a select number of genes including IL-1 $\beta$ , IL-8, MIP-1 $\beta$ , TNF, HME, Col-1, Col-3, Strom-1, and Strom-2 were compared with data collected from dot blot analysis. Results (not shown) were in close agreement and confirmed our earlier observations on the use of the microarray method for the quantitation of gene expression (3).

**Expression Profiles in Primary Chondrocytes and Synovio-cytes of Human RA Tissue.** Given the sensitivity and the specificity of this method, expression profiles of primary synovio-cytes and chondrocytes from diseased tissue were examined. Without prior exposure to inducing agents, low level expression of *c-jun*, G-CSF, IL-3, TNF- $\beta$ , MIF, and RANTES (regulated upon activation, normal T cell expressed and secreted) was seen as well as expression of MMPs, GelA, Strom-1, Col-1, and the three TIMPs. In this case, Col-2 hybridization was considered to be nonspecific because the second Col-2 target taken from the 3' end of the gene gave no

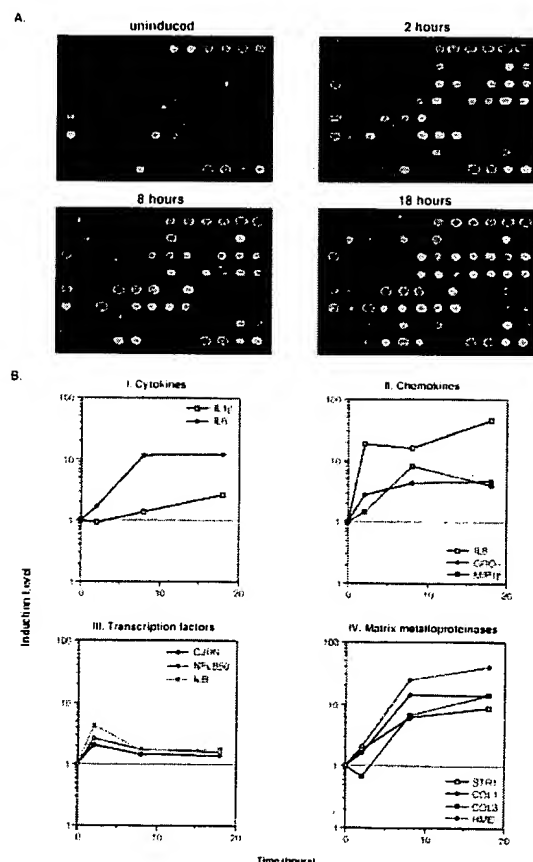


FIG. 3. Time course for IL-1 $\beta$  and TNF-induced SW1353 cells using the inflammation array (Fig. 1). (A) Pseudocolor representation of fluorescent scans correspond to gene expression levels at each time point. (B I–IV) Relative levels of selected genes at different time points compared with time zero.

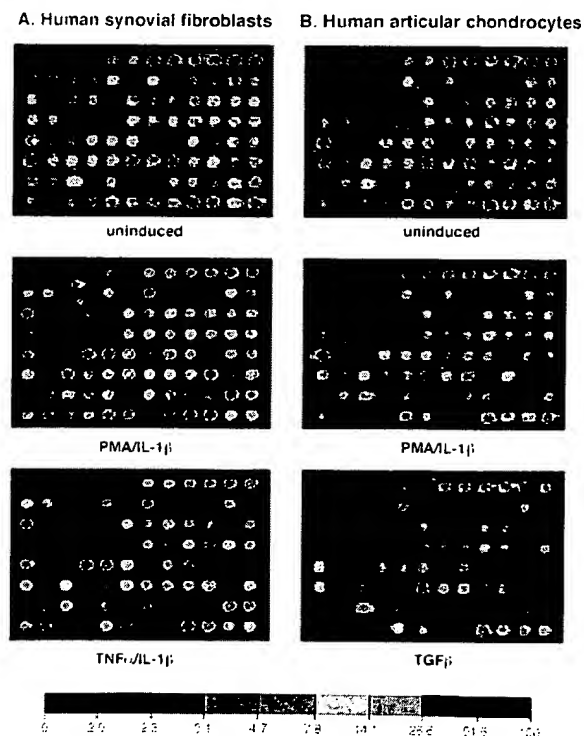


FIG. 4. Expression profiles for early passage primary synovio-cytes and chondrocytes isolated from RA tissue, cultured in the presence of 10% fetal calf serum and activated with PMA and IL-1 $\beta$ , or TNF and IL-1 $\beta$ , or TGF- $\beta$  for 18 hr. The color bars provide a comparative calibration scale between arrays and are derived from the *Arabidopsis* mRNA samples that are introduced in equal amounts during probe preparation.

signal. Treatment more so with PMA and IL-1, than TNF and IL-1, produced a dramatic up-regulation in expression of several genes in both of these primary cell types. These genes are as follows: the cytokine IL-6, the chemokines IL-8 and Gro-1 $\alpha$ , and the MMPs; Strom-1, Col-1, Col-3, and HME; and the adhesion molecule, vascular cell adhesion molecule 1 (VCAM-1). The surprise again is HME expression in these primary cells, for reasons discussed above. From these results, the expression profiles of synoviocytes and the chondrocytes appear very similar; the differences are more quantitative than qualitative. Treatment of the primary chondrocytes with the anabolic growth factor TGF- $\beta$  had an interesting profile in that it produced a remarkable down-regulation of genes expressed in both the untreated and induced state (Fig. 4).

Given the demonstrated effectiveness of this technology, a comparative analysis of two different inflammatory disease states was conducted with probes made from RA tissue and IBD samples. RA samples were from late stage rheumatoid synovial tissue, and IBD specimens were obtained from inflamed lower intestinal mucosa of patients with Crohn disease. With both the 96-element known gene microarray and the 1000-gene microarray of cDNAs selected from a peripheral human blood cell library (3), distinct differences in gene expression patterns were evident. On the 96-gene array, RA tissue samples from different affected individuals gave similar profiles (data not shown) as did different samples from the same individual (Fig. 5). These patterns were notably similar to those observed with primary synoviocytes and chondrocytes (Fig. 4). Included in the list of prominently up-regulated genes are IL-6, the MMPs Strom-1, Col-1, GelA, HME, and in

certain samples PUMP, TIMPs, particularly TIMP-1 and TIMP-3, and the adhesion molecule VCAM. Discernible levels of macrophage chemotactic protein 1 (MCP-1), MIF and RANTES were also noted. IBD samples were in comparison, rather subdued although IL-1 converting enzyme (ICE), TIMP-1, and MIF were notable in all the three different IBD samples examined here. In IBD-A, one of three individual samples, ICE, VCAM, Gro $\alpha$ , and MMP expression was more pronounced than in the others.

We also made use of a peripheral blood cDNA library (3) to identify genes expressed by lymphocytes infiltrating the inflamed tissues from the circulating blood. With the 1046-element array of randomly selected cDNAs from this library, probes made from RA and IBD samples showed hybridizations to a large number of genes. Of these, many were common between the two disease tissues while others were differentially expressed (data not shown). A complete survey of these genes was beyond the scope of this study, but for this report we picked three genes that were up-regulated in the RA tissue relative to IBD. These cDNAs were sequenced and identified by comparison to the GenBank database. They are TIMP-1, apoferritin light chain, and manganese superoxide dismutase (MnSOD). Differential expression of MnSOD was only observed in samples of RA tissue explants maintained in growth medium without serum for anywhere between 2 to 16 hr. These results also indicate that the expression profile of genes can be altered when explants are transferred to culture conditions.

## DISCUSSION

The speed, ease, and feasibility of simultaneously monitoring differential expression of hundreds of genes with the cDNA microarray based system (1-3) is demonstrated here in the analysis of a complex disease such as RA. Many different cell types in the RA tissue; macrophages, lymphocytes, plasma cells, neutrophils, synoviocytes, chondrocytes, etc. are known to contribute to the development of the disease with the expression of gene products known to be proinflammatory. They include the cytokines, chemokines, growth factors, MMPs, eicosanoids, and others (7, 11-14), and the design of the 96-element known gene microarray was based on this knowledge and depended on the availability of the genes. The technology was validated by confirming earlier observations on the expression of TNF by the monocyte cell line MM6, and of Col-1 and Col-3 expression in the chondrosarcoma cells and articular chondrocytes (9, 12). In our time-dependent survey the chronological order of gene activities in and between gene families was compared and the results have provided unprecedented profiles of the cytokines (TNF, IL-1, IL-6, GCSF, and MIF), chemokines (MIP-1 $\alpha$ , MIP-1 $\beta$ , IL-8, and Gro-1), certain transcription factors, and the matrix metalloproteinases (GelA, Strom-1, Col-1, Col-3, HME) in the macrophage cell line MM6 and in the SW1353 chondrosarcoma cells.

Earlier reports of cytokine production in the diseased state had established a model in which TNF is a major participant in RA. Its expression reportedly preceded that of the other cytokines and effector molecules (4). Our results strongly support these results as demonstrated in the time course of the MM6 cells where TNF induction preceded that of IL-1 $\alpha$  and IL- $\beta$  followed by IL-6 and GCSF. These expression profiles demonstrate the utility of the microarrays in determining the hierarchy of signaling events.

In the SW1353 chondrosarcoma cells, all the known MMPs and TIMPs were examined simultaneously. HME expression was discovered, which previously had been observed in only the stromal cells and alveolar macrophages of smoker's lungs and in placental tissue. Its presence in cells of the RA tissue is meaningful because its activity can cause significant destruction of elastin and basement membrane components (16, 17). Expression profiles of synovial fibroblasts and articular chondrocytes were remarkably similar and not too different from the SW1353 cells, indicating that the fibroblast and the chondrocyte can play equally aggressive roles in joint erosion. Prominent genes expressed were

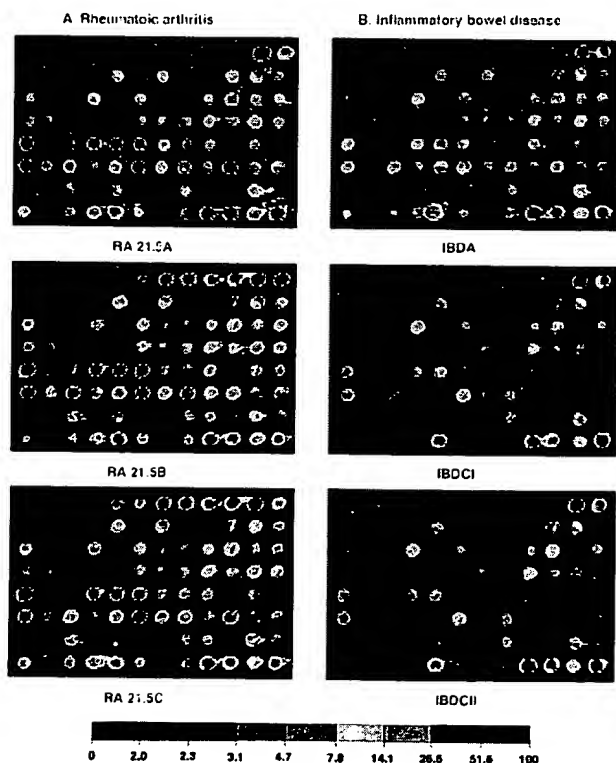


FIG. 5. Expression profiles of RA tissue (A) and IBD tissue (B). mRNA from RA tissue samples obtained from the same individual was isolated directly after excision (RA 21.5A) or maintained in culture without serum for 2 hr (RA 21.5B) or for 6 hr (RA 21.5C). Profiles from tissue samples of two other individuals (data not shown) were remarkably similar to the ones shown here. IBD-A and IBD-CI are from mRNA samples prepared directly after surgery from two separate individuals. For the IBD-CII probe, the tissue sample was cultured in medium without serum for 2 hr before mRNA preparation.

the MMPs, but chemokines and cytokines were also produced by these cells. The effect of the anabolic growth factor TGF- $\beta$  was profoundly evident in demonstrating the down regulation of these catabolic activities.

RA tissue samples undeniably reflected profiles similar to the cell types examined. Active genes observed were IL-3, IL-6, ICE, the MMPs including HME and TIMPs, chemokines IL-8, Gro $\alpha$ , MIP, MIF, and RANTES, and the adhesion molecule VCAM. Of the growth factors, fibroblast growth factor  $\beta$  was observed most frequently. In comparison, the expression patterns in the other inflammatory state (i.e., IBD) were not as marked as in the RA samples, at least as obtained from the tissue samples selected for this study.

As an alternative approach, the 1046 cDNA microarray of randomly selected genes from a lymphocyte library was used to identify genes expressed in RA tissue (3). Many genes on this array hybridized with probes made from both RA and IBD tissue samples. The results are not surprising because inflammatory tissue is abundantly supplied with cell types infiltrating from the circulating blood, made apparent also by the high levels of chemokine expression in RA tissue. Because of the magnitude of the effort required to identify all the hybridized genes, we have for this report chosen to describe only three differentially expressed genes mainly to verify this method of analysis.

Of the large number of genes observed here, a fair number were already known as active participants in inflammatory disease. These are TNF, IL-1, IL-6, IL-8, GCSF, RANTES, and VCAM. The novel participants not previously reported are HME, IL-3, ICE, and Gro $\alpha$ . With our discovery of HME expression in RA, this gene becomes a target for drug intervention. ICE is a cysteine protease well known for its IL-1 $\beta$  processing activity (18), and recognized for its role in apoptotic cell death (19). Its expression in RA tissue is intriguing. IL-3 is recognized for its growth-promoting activity in hematopoietic cell lineages, is a product of activated T cells (20), and its expression in synovocytes and chondrocytes of RA tissue is a novel observation.

Like IL-8, Gro $\alpha$ , is a C-X-C subgroup chemokine and is a potent neutrophil and basophil chemoattractant. It down-regulates the expression of types I and III interstitial collagens (21, 22) and is seen here produced by the MM6 cells, in primary synovocytes, and in RA tissue. With the presence of RANTES, MCP, and MIP-1 $\beta$ , the C-C chemokines (23) migration and infiltration of monocytes, particularly T cells, into the tissue is also enhanced (5) and aid in the trafficking and recruitment of leukocytes into the RA tissue. Their activation, phagocytosis, degranulation, and respiratory bursts could be responsible for the induction of MnSOD in RA. MnSOD is also induced by TNF and IL-1 and serves a protective function against oxidative damage. The induction of the ferritin light chain encoding gene in this tissue may be for reasons similar to those for MnSOD. Ferritin is the major intracellular iron storage protein and it is responsive to intracellular oxidative stress and reactive oxygen intermediates generated during inflammation (24, 25). The active expression of TIMP-1 in RA tissue, as detected by the 1000-element array, is no surprise because our results have repeatedly shown TIMP-1 to be expressed in the constitutive and induced states of RA cells and tissues.

The suitability of the cDNA microarray technology for profiling diseases and for identifying disease related genes is well documented here. This technology could provide new

targets for drug development and disease therapies, and in doing so allow for improved treatment of chronic diseases that are challenging because of their complexity.

We would like to thank the following individuals for their help in obtaining reagents or providing cDNA clones to use as templates in target preparation: N. Arai, P. Cannon, D. R. Cohen, T. Curran, V. Dixit, D. A. Geller, G. I. Goldberg, M. Karin, M. Lotz, L. Matrisian, G. Nolan, C. Lopez-Otin, T. Schall, S. Shapiro, I. Verma, and H. Van Wart. Support for R.W.D., M.S., and R.A.H. was provided by the National Institutes of Health (Grants R37HG00198 and HG00205).

1. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* 270, 467-470.
2. Shalon, D., Smith, S. & Brown, P. O. (1996) *Genome Res.* 6, 639-645.
3. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. & Davis, R. W. (1996) *Proc. Natl. Acad. Sci. USA* 93, 10614-10619.
4. Feldmann, M., Brennan, F. M. & Maini, R. N. (1996) *Rheumatoid Arthritis Cell* 85, 307-310.
5. Schall, T. J. (1994) in *The Cytokine Handbook*, ed. Thomson, A. W. (Academic, New York), 2nd Ed., pp. 410-460.
6. Lotz, M. F., Blanco, J., Von Kempis, J., Dudley, J., Maier, R., Villiger, P. M. & Geng, Y. (1995) *J. Rheumatol.* 22, Supplement 43, 104-108.
7. Birkedal-Hansen, H., Moore, W. G. I., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., DeCarlo, A. & Engler, J. A. (1993) *Crit. Rev. Oral Biol. Med.* 4, 197-250.
8. Zeigler-Heitbrock, H. W. L., Thiel, E., Futterer, A., Volker, H., Wirtz, A. & Reithmüller, G. (1988) *Int. J. Cancer* 41, 456-461.
9. Borden, P., Solymar, D., Sucharczuk, A., Lindman, B., Cannon, P. & Heller, R. A. (1996) *J. Biol. Chem.* 271, 23577-23581.
10. Gader, S. J. & Woolley, D. E. (1987) *Rheumatol. Int.* 7, 13-22.
11. Harris, E. D., Jr. (1990) *New Engl. J. Med.* 322, 1277-1289.
12. Firestein, G. S. (1996) in *Textbook of Rheumatology*, eds. Kelly, W. N., Harris, E. D., Ruddy, S. & Sledge, C. B. (Saunders, Philadelphia), 5th Ed. pp. 5001-5047.
13. Alvaro-Garcia, J. M., Zvaifler, Nathan J., Brown, C. B., Kaushansky, K. & Firestein, Gary S. (1991) *J. Immunol.* 146, 3365-3371.
14. Firestein, G. S., Alvaro-Garcia, J. M. & Maki, R. (1990) *J. Immunol.* 144, 3347-3352.
15. Pradines-Figueres, A. & Raetz, C. R. H. (1992) *J. Biol. Chem.* 267, 23261-23268.
16. Shapiro, S. D., Kobayashi, D. L. & Ley, T. J. (1993) *J. Biol. Chem.* 268, 23824-23829.
17. Shipley, M. J., Wesselschmidt, R. L., Kobayashi, D. K., Ley, T. J. & Shapiro, S. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 3042-3046.
18. Cerretti, D. P., Kozlosky, C. J., Mosley, B., Nelson, N., Van Ness, K., Greenstreet, T. A., March, C. J., Kronheim, S. R., Druck, T., Cannizaro, L. A., Huebner, K. & Black, R. A. (1992) *Science* 256, 97-100.
19. Miura, M., Zhu, H., Rotello, R., Hartweg, E. A. & Yuan, J. (1993) *Cell* 75, 653-660.
20. Arai, K., Lee, F., Miyajima, A., Shoichiro, M., Arai, N. & Takashi, Y. (1990) *Annu. Rev. Biochem.* 59, 783-836.
21. Geiser, T., Dewald, B., Ehrenguber, M. U., Lewis, I. C. & Baggiolini, M. (1993) *J. Biol. Chem.* 268, 15419-15424.
22. Unemori, E. N., Amento, E. P., Bauer, E. A. & Horuk, R. (1993) *J. Biol. Chem.* 268, 1338-1342.
23. Robinson, E., Keystone, E. C., Schall, T. J., Gillet, N. & Fish, E. N. (1995) *Clin. Exp. Immunol.* 101, 398-407.
24. Roeser, H. (1980) in *Iron Metabolism in Biochemistry and Medicine*, eds. Jacobs, A. & Worwood, M. (Academic, New York), Vol. 2, pp. 605-640.
25. Kwak, E. L., Larochelle, D. A., Beaumont, C., Torti, S. V. & Torti, F. M. (1995) *J. Biol. Chem.* 270, 15285-15293.

FOCUS - 17 of 19 DOCUMENTS

Copyright 1997 PR Newswire Association, Inc.  
PR Newswire

August 11, 1997, Monday

SECTION: Financial News

DISTRIBUTION: TO BUSINESS AND MEDICAL EDITORS

LENGTH: 478 words

HEADLINE: Eli Lilly & Co. and Acacia Biosciences Enter Into Research Collaboration;  
First Corporate Agreement for Acacia's Genome Reporter Matrix(TM)

DATELINE: RICHMOND, Calif., Aug. 11

BODY:

Acacia Biosciences and Eli Lilly and Company (Lilly) announced today the signing of a joint research collaboration to utilize Acacia's Genome Reporter Matrix(TM) (GRM) to aid in the selection and optimization of lead compounds. Under the collaboration, Acacia will provide chemical and biological profiles on a class of Lilly's compounds for an undisclosed fee.

Acacia's GRM is an assay-based computer modeling system that uses yeast as a miniature ecosystem. The GRM can profile the extent, nature and quantity of any changes in gene expression. Because of the similarities between the yeast and human genome, the system serves as an excellent surrogate for the human body, mimicking the effects induced by a biologically active molecule.

"Using yeast as a model organism for lead optimization makes a lot of sense given the high degree of homology with human metabolic pathways," said William Current of Lilly Research Laboratories. "Acacia's innovative GRM has the potential to provide enormous insight into the therapeutic impact of our compounds and make the drug discovery process more rational. It should substantially accelerate the development process."

"This first agreement with a major pharmaceutical company is an important milestone in the development of Acacia," said Bruce Cohen, President and CEO of Acacia. "The deal is in line with our strategy of establishing alliances that will allow our collaborators to use genomic profiles to identify and optimize compounds within their existing portfolios. In the long run, this technology can be used to characterize large scale combinatorial libraries, predict side effects prior to clinical trials and resurrect drugs that have failed during clinical trials."

The GRM incorporates two critical elements: chemical response profiles and genetic response profiles. The chemical response profiles measure the change in gene expression caused by potential therapeutics and then rank genes with altered expressions by degree of response. The genetic response profiles measure changes in gene expression caused by mutations in the genes encoding potential targets of pharmaceuticals; these genetic response profiles represent gold standards in drug discovery by defining the response profile expected for drugs with perfect selectivity and specificity. By comparing the two profiles, one can analyze a potential drug candidate's ability to mimic the action of a 'perfect' drug.

Acacia Biosciences is a functional genomics company developing proprietary technologies to enhance the speed and efficacy of drug discovery and development. Acacia's Genome Reporter Matrix capitalizes on the latest advances in genomics and combinatorial chemistry to generate comprehensive profiles of drug candidates' in vivo activity.

SOURCE Acacia Biosciences  
CONTACT: Bruce Cohen, President and CEO of Acacia Biosciences, 510-669-2330 ext. 103 or Media: Linda Seaton of Feinstein

LOAD-DATE: August 12, 1997

**The Bioreactor Market: Steady Growth Expected**

The worldwide market for all bioreactors was valued at \$275 million for 1997, and is expected to be worth \$380 million by 2002.

Types: stirred-tank, gas-lift, etc.

W1 GE281N  
V.17 NO.16  
C.01  
1997  
SEQ: G04575000  
TI: GENETIC ENGINEERING NEWS

# GENETIC ENGINEERING

## GEN NEWS

BIOTECHNOLOGY

08/25/97

BIOPROCESS

BIORESEARCH • TECHNOLOGY TRANSFER

### Contents

European Patent Standards Moving	4
Bioprocess Development Packages	13
Trends in Manufacturing Development	14
GC/MS for Drug Bioprocess	16
Advances in Electroporation	19
New Products	21
New Drug Clinical Drug Discovery	27
Corporate Profile: Pangea Systems	28
Channel Watch	29
European Funding	30
Wall Street Outlook	31
Case Study	32
Trade Industry	33
Channel Watch	37
Case Study	40
Case Study	41
Case Study	41
Case Study	42

## Pharmagene Raises More Capital for Research on Human Tissues

By Sophia Fox

**P**harmagene, the Royston, U.K.-based biopharmaceutical company specializing in the use of human biomaterials for drug discovery research, has raised a further £5 million from a group of investors led by 3i and Abacus Nominees. The funding will enable the company to expand both its human biomaterials collection and its capabilities across a range of proprietary platform technologies.

Gordon Baxter, Ph.D., Pharmagene's cofounder and chief operating officer, claimed, "by the end of this year Pharmagene will have access to the largest collection of human RNAs and proteins anywhere in the world, and a range of innovative, yet robust technologies

SEE PHARMAGENE, P. 9

## Perkin-Elmer Acquires PerSeptive to Expand Its Capabilities in Gene-Based Drug Discovery

By John Sterling

**P**erkin-Elmer's (PE; Norwalk, CT) decision last month to acquire PerSeptive Biosystems (Frammingham, MA) via a \$360 million stock swap was designed to strengthen PE in terms of broad capabilities in gene-based drug discovery. The company's main goal is to develop new products to improve the integration of genetic and protein research.

"This merger will enhance our position as an effective provider of innovative, integrated platforms enabling our customers to be more efficient and cost-effective in bringing new pharmaceuticals to market," says Tony L. White, PE's chairman, president and CEO. "The combination of our two companies should bolster our presence in the life sciences, [and it is our] belief that we must take bold action now to lead the emerging era of molecular medicine with leading positions in both genetic and protein analysis."

A driving force behind the merger is the vast amount of genet-



Perkin-Elmer acquired PerSeptive Biosystems for \$360 million to obtain new technologies in mass spectrometry, bioseparations and purification for product development projects, spanning the range from genomics to proteomics.

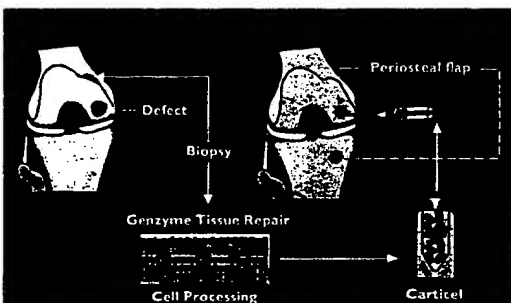
ic information about human disease that is being accumulated by researchers and biotech companies working in the area of genomics. It is becoming increasingly obvious that these data need to be complemented with technologies for

studying proteins and protein networks—a field known as proteomics (see GEN, September 1, 1997, p. 1).

PE officials, who claim that MALDI-TOF (Matrix Assisted Laser Desorption/Ionization) technology is a key to

SEE ACQUISITION, P. 10

## FDA OKs Genzyme's Carticel Product for Damage to Knees



Carticel, which was approved for the repair of clinically significant, symptomatic cartilaginous defects of the femoral condyle (medial, lateral or trochlear) caused by acute or repetitive trauma, employs a proprietary process to grow autologous cartilage cells for implantation.

By Naomi Pfeiffer

**T**he FDA has approved a knee-cartilage replacement product made by Genzyme Tissue Repair (Cambridge, MA), a tracking-stock division of Genzyme Corp., for people with trauma-damaged knees.

Carticel™ (autologous cultured chondrocytes) is the first product to be licensed under the FDA's pro-

SEE GENZYME, P. 8

## Strategies for Target Validation Streamline Evaluation of Leads

By Vicki Glaser

**A**acacia Biosciences (Richmond, CA) last month announced its first agreement with a major pharmaceutical company, signing a deal with Eli Lilly (Indianapolis, IN) to use Acacia's Genome Reporter Matrix (GRM) to select and optimize some of Lilly's lead compounds. Acacia's yeast-based system for profiling drug activity is useful for evaluating the therapeutic potential of lead compounds, and it also has a role in the identification and validation of new drug targets.

"We're using the ecosystem of a cell to allow us to deduce the mechanism of action and target for any chemical," explains Bruce Cohen, president and CEO. "We screen for every target in a cell simultaneously...using transcription as a readout

for how a cell is adapting to any perturbation," he says.

The GRM technology consists of two main databases: one is the genetic response profile, showing the effects of mutations in each individual yeast gene and compensatory gene regulatory mechanisms; the other is the chemical response profile, which documents changes in gene expression in response to chemical compounds. Computational analysis and pattern matching between the genetic and chemical profiles yields information on the specificity, potency and side-effects risk of a drug lead.

### Targeting Targets

No longer is mapping and sequencing a gene—or the human genome—an end unto itself, but

SEE TARGET, P. 15

## Sticky Ends

Avigen received two grants from the NIH & University of California for research on gene therapy for treatment of cancer & HIV infections...MRL Pharmaceutical Services, of Reston, VA, launched the TSN Bug Finder, which is able to locate & retrieve client-specified microorganisms in real-time...GenSia Sidor, Inc. will move its corporate staff from San Diego to Irvine, CA, by end of year...

FDA accepted NDA from Sepracor for levalbuterol HCl inhalation solution...An \$11.7M mezzanine financing has been closed by Activated Cell Therapy, which changed its name to Dendreon Corporation...Astra AB will build major research facility in Waltham, MA, and is also relocating Astra research facility from Rochester to Boston area...Prolifix Ltd. team used a small peptide to inhibit the E2F protein complex and induced

apoptosis in mammalian tumor cells...Vertex Pharmaceuticals, Inc. and Alpha Therapeutic Corp. ended an agreement to develop VX-366 for treatment of inherited hemoglobin disorders...Navicore received Phase I SBIR grant for up to \$100,000 from NIH for development of prototype of its Naviflow technology for high-throughput screening...Covance Inc. will invest \$21 million in expansion and renovation of its facility in Indianapolis, IN.



## Target

from page 1

merely a means to an end. The critical next step is to validate the gene and its protein product as a potential drug target. The Human Genome Project continues to produce a treasure chest of expressed sequence tags (ESTs) and a tantalizing array of complete gene sequences.

Companies are applying a variety of functional genomic strategies to link genes to specific diseases and to multigenic phenotypes. Yet the ultimate challenge for pharmaceutical companies is to sift through all the sequence and differential gene expression data to identify the best targets for drug discovery.

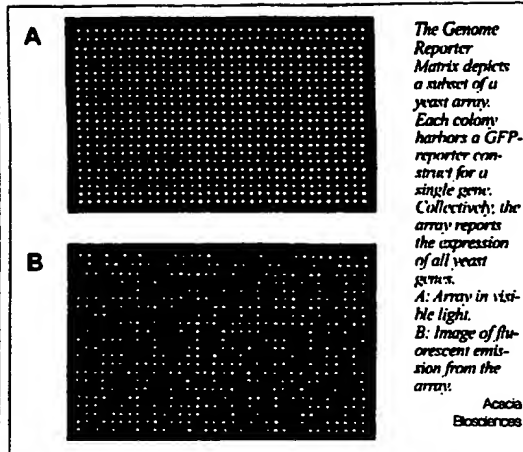
Spinning off technology developed at the University of North Carolina (Chapel Hill), Cytogen Corp. (Princeton, NJ) formed its wholly owned subsidiary AxCell Biosciences earlier this year. The young company is building a protein interaction database, cataloging all the interactions the modular domains of proteins can engage in with a

range of ligands, in order to gain insight into protein function and to select the most critical interaction to target for drug development.

AxCell's cloning-of-ligand-targets (COLT) technology employs "recognition units" from the company's genetic diversity library (GDL) to map functional protein interactions and quantitate their affinity. The company's inter-functional proteomic database (IFP-dbase) elucidates protein interaction networks and structure-activity relationships based on ligand affinity with protein modular domains.

### Defining Disease Pathways

Signal Pharmaceuticals, Inc.'s (San Diego, CA) integrated drug target and discovery effort is based on mapping gene-regulating pathways in cells and identifying small molecules that regulate the activation of those genes. In collaboration with academic researchers, the company has identified a large number of regulatory proteins in several mitogen-activated protein (MAP) kinase pathways (including the JNK, FRK and p38



**The Genome Reporter Matrix depicts a subset of a yeast array. Each colony harbors a GFP-reporter construct for a single gene. Collectively, the array reports the expression of all yeast genes. A: Array in visible light. B: Image of fluorescence emission from the array.**

Acacia Biosciences

signaling pathways), which Signal is evaluating for the treatment of autoimmune, inflammatory, cardiovascular and neurologic diseases, and cancer. Other target identification

programs focus on the NF- $\kappa$ B pathway, estrogen-related genes and central/peripheral nervous system genes.

Regulating cytokine production in immune and inflammatory disorders,

and modifying bone metabolism to treat osteoporosis are the focus of Signal's collaboration with Tanabe Sanyaku (Osaka, Japan). Signal has partnered with Organon/Akzo Nobel (Netherlands) to identify estrogen-responsive genes as targets for treating neurodegenerative and psychiatric diseases, atherosclerosis and ischemia, and with Roche Bioscience (Palo Alto, CA) to develop human peripheral nerve cell lines for the discovery of treatments for pain and incontinence.

Exelixis' (S. San Francisco, CA) strategy for target selection is to define disease pathways and identify regulatory molecules that activate or inhibit those biochemical/genetic pathways. Based on the finding that these pathways are conserved across species, the company is studying the model genetic systems of *Drosophila* and *Caenorhabditis elegans*. Using its PathFinder technology, Exelixis systematically introduces mutations into the genomes of these model organisms, looking for mutations that enhance or suppress the target disease-related gene. These novel genes then become the basis of drug screening assays.

Cadus Pharmaceutical Corp. (Tarrytown, NY) is identifying surrogate ligands to newly discovered orphan G-protein coupled transmembrane receptors of unknown function to determine the suitability of the receptors as drug targets. Inserting the novel receptor in a yeast system yields a ligand that activates the receptor. Access to a surrogate ligand allows the company to screen for receptor antagonists in the yeast system.

"The antagonist plus the surrogate ligand gives you two probes—an on probe and an off probe—which allows you to look at function," explains David Webb, Ph.D., vp of research and chief scientific officer. A surrogate ligand also provides information on which G-protein interacts with the orphan receptor and its associated signaling pathways, further clarifying the role of the receptor as a potential drug target. Cadus' collaboration with SmithKline (Philadelphia) capitalizes on Cadus' ability to determine orphan receptor function, applying the technology to SmithKline's proprietary, newly discovered G-protein receptors.

Cadus' recombinant yeast system can also be used to screen cell and tissue extracts for natural ligands, and the company is accelerating its internal drug-discovery efforts in the areas of cancer, inflammation and allergy. A recent equity investment in Axiom Biotechnologies (San Diego, CA) gave Cadus a license to Axiom's high-throughput pharmacologic screening system for lead optimization and discovery.

As its name implies, geneNetworks (Alameda, CA) focuses on identifying gene networks that contribute to multigenic phenotypes and complex disease processes. The integration of mouse and human genetic studies forms the basis of the technology. The Genome Tagged Mice database in development will serve as a library of natural mouse genetic and phenotypic variation. Disease-related genes identified in mice are then evaluated in human family- and population-based studies to confirm their clinical relevance and linkages to pathophysiologic traits.

### Blocking Gene Expression

Inactivating a gene known to be expressed in association with a particular disease is one approach to identifying appropriate therapeutic targets. The target validation and discovery program at Ribozyme Pharmaceuticals, Inc. (Boulder, CO) applies the company's ribozyme technology to achieve selective inhibition of gene expression in cell culture and in animals.

Correlation of the gene expression inhibition with phenotype can

SEE TARGET, P. 38

## A strong chemical combination to help you grow. And flourish.

Three hundred million dollars and ten years of hard work. That's what it costs to bring your biotechnology-derived therapeutic to the marketplace.

Which means, no room for error.

Which means, in turn, you'd be wise to tap into the combined capabilities of Mallinckrodt and J.T.Baker: dual sources, trusted names for your chemical raw materials.

Two separate GMP-produced brands offering the control of a single quality system and the convenience of a single audit process.

We offer comprehensive product lines including USP salts, bioreagents, high purity solvents and chromatography products in Beaker to Bulk™ packaging for easy scale-up.

Call 1-800-582-2537, or access our website at <http://www.mallinckrodt.com>. For dual chemical sources dedicated to helping you grow. Flourish. Succeed!

MALLINCKRODT





## Target

from page 15

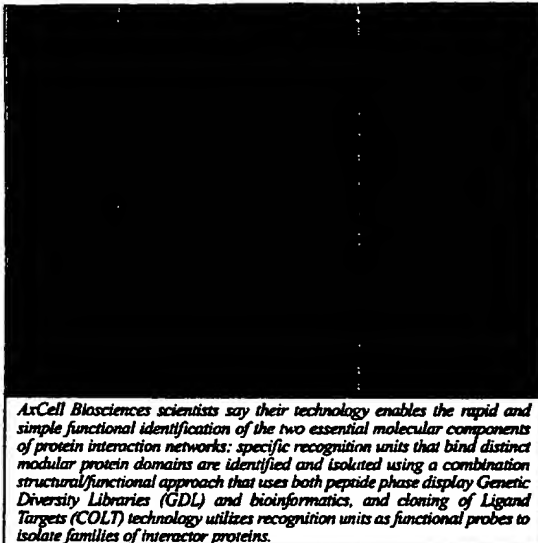
suggest the relative importance of the gene in disease pathology. The company's nuclease-resistant ribozymes form the basis of a collaboration with Schering AG (Germany) for drug target validation and the development of ribozyme-based therapeutic agents, and with Chiron Corp. (Emeryville, CA) for target validation.

With several antisense compounds now progressing through clinical trials, the concept of using oligonucleotides to inhibit gene activity is not new. But rather than focusing on therapeutics development, Sequitur, Inc. (Natick, MA) is creating antisense compounds for the purpose of determining gene function and validating drug targets. Clients typically provide the one-year-old company with the sequence (or EST) of a potential gene target and, in return, Sequitur custom designs a series of three to six antisense compounds that yield a three-to-ten-fold inhibition of the target gene in cell culture. The company also provides oligofectins, a series of cationic lipids, to deliver the oligonucleotides to a variety of cultured cells.

"Differential expression information is just for correlation, it doesn't tell function or confirm what would be a good target," says Tod Woolf, Ph.D., director of technology development at Sequitur. Whereas, antisense compounds will inhibit a target, Sequitur offers both phosphorothioate DNA antisense compounds, and its proprietary Next Generation chimeric oligonucleotides, which have a higher hybridization affinity, greater specificity and reduced toxicity, according to the company.

### Mining Pathogen Genomes

Companies such as Human Genome Sciences (HGS; Rockville, MD), Incyte (Palo Alto, CA),



*ArCell Biosciences scientists say their technology enables the rapid and simple functional identification of the two essential molecular components of protein interaction networks: specific recognition units that bind distinct modular protein domains are identified and isolated using a combination structural/functional approach that uses both peptide phase display Genomic Diversity Libraries (GDL) and bioinformatics, and cloning of Ligand Targets (COLT) technology utilizes recognition units as functional probes to isolate families of interactor proteins.*

Millennium Pharmaceuticals Inc. (Cambridge, MA) and Genome Therapeutics (Waltham, MA) are relying on high-speed DNA sequencing, positional cloning and other strategies to identify specific microbial genomic sites that would be good targets for infectious disease therapeutics.

HGS recently completed sequencing of the bacterial pathogen *Streptococcus pneumoniae*, which is the focus of an agreement with Hoffmann-La Roche (Basel, Switzerland). Roche will use the sequence data to develop new anti-infectives against *S. pneumoniae*. HGS and Roche have expanded their collaboration to include a nonexclusive license to access sequence information for the intestinal bacterium *Enterococcus faecalis*.

Incyte Pharmaceuticals has completed one-fold coverage of the *Candida albicans* genome, identifying

60% of the genes of this fungal pathogen. This genome will become part of the company's PathoSeq microbial database. Incyte recently introduced the ZooSeq animal gene sequence and expression database. The database will provide genomic information across various species commonly used in preclinical drug testing, which may help to better define potential drug targets.

Millennium Pharmaceuticals continues to report success in identifying novel drug targets, having recently discovered a novel chemokine called neurotactin and a new class of MAD-related proteins that inhibit transforming growth factor beta (TGF- $\beta$ ) signaling. The company also received U.S. patent coverage for the tub genes, believed to play a role in obesity, and for the gene that encodes the protein melanastatin, which appears to suppress metastasis in malignant melanoma.

## Pangea

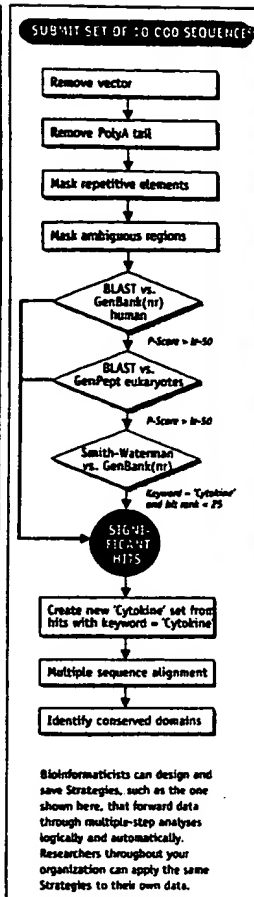
from page 28

Smith, now a computer programmer, is an expert in systems integration, Internet technologies and the application of industrial engineering principles to the drug discovery process. Before co-founding Pangea, he was the manager of software development at Attorney's Briefcase, a legal research software company.

By being "in the trenches" with customers and collaborators, Bellenson and Smith sensed the frustration of pharmaceutical researchers whose incompatible tools have impeded their progress. According to Bellenson, "Most of them are geared toward analyzing one molecule at a time. It's like emptying the ocean with an eye dropper—an incompatible eye dropper at that. A pharmaceutical company may have 30 different drug discovery teams with various approaches. The problem is to manage the process of experimenting with a lot of different approaches, to automate while maintaining flexibility."

GeneWorld 2.1 enables "integration of the entire target discovery and validation process," Bellenson says. The commercial software package coordinates the entire process of sequence-data analysis and can be integrated with other programs and databases, according to Smith, who adds that it handles thousands of sequence results, organizes and automates annotation and seamlessly interacts with growing genome databases. Simple forms and menus enable users to turn raw sequence data into crucial knowledge for drug discovery by applying algorithms to sequences, creating custom analysis strategies and producing useful reports, without the need for writing computer code. GeneWorld 2.1 runs on a variety of platforms and operating systems.

Pairing industrial relational database-management systems with a web-browser interface, Pangea's Operating System of Drug Discovery is an open-computing framework that allows client/server and Java-enabled web-based technologies to collect, organize and analyze drug discovery information for pharmaceutical companies to simplify and accelerate drug discovery. The technology unites automated genomics database analysis for drug target site selection, chemical information database analysis and large-scale combinatorial chemistry project management and high-throughput screening project management for drug lead efficacy analysis. Pangea officials maintain that these integrated elements provide a unified environment for chemists, biologists and others involved in the drug discovery process to work together with



commercial and public domain software.

Pangea's Operating System of Drug Discovery can accommodate Sybase, Oracle or Informix relational database-management systems and any version of UNIX. It absorbs new data formats, databases, algorithms and analysis paradigms into the automated workflow without software modifications. Netscape Navigator provides a friendly user interface from PC, Macintosh, and UNIX workstations.

In the near term, Pangea plans to complete its bioinformatics core with two more programs. Gene Foundry, a sample tracking and workflow sequence package for DNA sequence and fragment information, will also offer interaction with robots, reagent tracking and troubleshooting. Gene Thesaurus, the other package is a "warehouse of bioinformatics data," says Bellenson.

## Europe

from page 30

GTAC Chairman, Professor Norman C. Nevin, said 1996 saw "four important developments": an increase in enquiries and submissions made to GTAC; an increase in the complexity of submitted protocols; a continuing shift from gene therapy for single-gene disorders toward strategies aimed at tumour destruction in cancer; and a growth in international sponsorship of U.K. gene therapy trials.

Since 1993, GTAC and its predecessor, the Clotier Committee, have approved 18 U.K. gene therapy clinical trials (13 of which have been carried out), which are listed in the report. The disease areas targeted by these trials include severe combined immunodeficiency (1 trial), cystic fibrosis (6), metastatic melanoma (2), lymphoma (2), neuroblastoma (1), breast cancer (1), Hurler's syndrome (1), cervical cancer (1), glioblastoma

breast cancer, breast cancer with liver metastases, glioblastoma, malignant ascites due to gastrointestinal cancer and ovarian cancer.

Copies of the GTAC third annual report are available from the GTAC Secretariat, Wellington House, 133-155 Waterloo Road, London SE1 8UG, U.K.

### Coated Lenses Prevent PCO

Scientists in the U.K. say it may be possible to prevent posterior capsule opacification (PCO), a common complication following cataract surgery, by using the implanted polymethylmethacrylate (PMMA) intraocular lens as a drug delivery system. PCO occurs in 30-50% of cataract surgery patients as a result of stimulated cell growth within the remaining capsular bag. The condition causes a decline in visual acuity and requires expensive laser treatment, thus negating the routine use of cataract surgery in underdeveloped countries, explains G. Duncan, at the



## NEW HIGH SPECIFIC ACTIVITY MICROBIAL ALKALINE PHOSPHATASE from Biocatalysts

Biocatalysts Limited, the British speciality enzyme company, has developed a completely new type of alkaline phosphatase with many advantages over the types most commonly used.

It is of microbial origin with a high specific activity (unlike that from *E. coli*) and with higher temperature and storage stability compared to that from calf intestine.

This is the first of several new generation diagnostic enzymes being developed by Biocatalysts Limited with greatly improved stability.

- Non-animal source, no risk of BSE or animal virus contamination
- Higher temperature stability than calf intestine
- Much higher specific activity than from *E. coli*
- Very high storage stability even in the absence of glycerol

For further details on alkaline phosphatase and our other diagnostic enzymes contact us direct at the address below or within North America contact our US Distributor Kaltron-Pettibone 'phone: 630 350 1118 or fax: 630-350-1606

Biocatalysts Limited  
Treforest Industrial Estate Pontypridd Wales UK CF37 5UD  
Tel: +44 (0)1443 843712 Fax: +44 (0)1443 841214  
e-mail: Kelly@Biocatalysts.com



- Fischer-Vize, *Science* 270, 1828 (1995).
35. T. C. James and S. C. Elgin, *Mol. Cell Biol.* 6, 3862 (1986); R. Paro and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* 88, 263 (1991); B. Tschiersch et al., *EMBO J.* 13, 3822 (1994); M. T. Madiredi et al., *Cell* 87, 75 (1996); D. G. Stokes, K. D. Tartof, R. P. Perry, *Proc. Natl. Acad. Sci. U.S.A.* 93, 7137 (1996).
  36. P. M. Palosaari et al., *J. Biol. Chem.* 266, 10750 (1991); A. Schmitz, K. H. Gartemann, J. Fiedler, E. Grund, R. Eichenlaub, *Appl. Environ. Microbiol.* 58, 4068 (1992); V. Sharma, K. Suvama, R. Meganathan, M. E. Hudspeth, *J. Bacteriol.* 174, 5057 (1992); M. Kanazawa et al., *Enzyme Protein* 47, 9 (1993); Z. L. Boynton, G. N. Bennet, F. B. Rudolph, *J. Bacteriol.* 178, 3015 (1996).
  37. M. Ho et al., *Cell* 77, 859 (1994).
  38. W. Hendriks et al., *J. Cell Biochem.* 59, 418 (1995).
  39. We thank H. Skaletsky and F. Lewitter for help with

sequence analysis; Lawrence Livermore National Laboratory for the flow-sorted Y cosmid library; and P. Bain, A. Bortvin, A. de la Chapelle, G. Fink, K. Jegalian, T. Kawaguchi, E. Lander, H. Lodish, P. Matsudaira, D. Merke, U. RajBhandary, R. Rajio, S. Rozen, A. Schwartz, C. Sun, and C. Tifford for comments on the manuscript. Supported by NIH.

28 April 1997; accepted 9 September 1997

## Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale

Joseph L. DeRisi, Vishwanath R. Iyer, Patrick O. Brown\*

DNA microarrays containing virtually every gene of *Saccharomyces cerevisiae* were used to carry out a comprehensive investigation of the temporal program of gene expression accompanying the metabolic shift from fermentation to respiration. The expression profiles observed for genes with known metabolic functions pointed to features of the metabolic reprogramming that occur during the diauxic shift, and the expression patterns of many previously uncharacterized genes provided clues to their possible functions. The same DNA microarrays were also used to identify genes whose expression was affected by deletion of the transcriptional co-repressor *TUP1* or overexpression of the transcriptional activator *YAP1*. These results demonstrate the feasibility and utility of this approach to genomewide exploration of gene expression patterns.

The complete sequences of nearly a dozen microbial genomes are known, and in the next several years we expect to know the complete genome sequences of several metazoans, including the human genome. Defining the role of each gene in these genomes will be a formidable task, and understanding how the genome functions as a whole in the complex natural history of a living organism presents an even greater challenge.

Knowing when and where a gene is expressed often provides a strong clue as to its biological role. Conversely, the pattern of genes expressed in a cell can provide detailed information about its state. Although regulation of protein abundance in a cell is by no means accomplished solely by regulation of mRNA, virtually all differences in cell type or state are correlated with changes in the mRNA levels of many genes. This is fortuitous because the only specific reagent required to measure the abundance of the mRNA for a specific gene is a cDNA sequence. DNA microarrays, consisting of thousands of individual gene sequences printed in a high-density array on a glass microscope slide (1, 2), provide a practical and economical tool for studying gene expression on a very large scale (3–6).

*Saccharomyces cerevisiae* is an especially

favorable organism in which to conduct a systematic investigation of gene expression. The genes are easy to recognize in the genome sequence, cis regulatory elements are generally compact and close to the transcription units, much is already known about its genetic regulatory mechanisms, and a powerful set of tools is available for its analysis.

A recurring cycle in the natural history of yeast involves a shift from anaerobic (fermentation) to aerobic (respiration) metabolism. Inoculation of yeast into a medium rich in sugar is followed by rapid growth fueled by fermentation, with the production of ethanol. When the fermentable sugar is exhausted, the yeast cells turn to ethanol as a carbon source for aerobic growth. This switch from anaerobic growth to aerobic respiration upon depletion of glucose, referred to as the diauxic shift, is correlated with widespread changes in the expression of genes involved in fundamental cellular processes such as carbon metabolism, protein synthesis, and carbohydrate storage (7). We used DNA microarrays to characterize the changes in gene expression that take place during this process for nearly the entire genome, and to investigate the genetic circuitry that regulates and executes this program.

Yeast open reading frames (ORFs) were amplified by the polymerase chain reaction (PCR), with a commercially available set of primer pairs (8). DNA microarrays, containing approximately 6400 distinct DNA sequences, were printed onto glass slides by

using a simple robotic printing device (9). Cells from an exponentially growing culture of yeast were inoculated into fresh medium and grown at 30°C for 21 hours. After an initial 9 hours of growth, samples were harvested at seven successive 2-hour intervals, and mRNA was isolated (10). Fluorescently labeled cDNA was prepared by reverse transcription in the presence of Cy3 (green)- or Cy5 (red)-labeled deoxyuridine triphosphate (dUTP) (11) and then hybridized to the microarrays (12). To maximize the reliability with which changes in expression levels could be discerned, we labeled cDNA prepared from cells at each successive time point with Cy5, then mixed it with a Cy3-labeled "reference" cDNA sample prepared from cells harvested at the first interval after inoculation. In this experimental design, the relative fluorescence intensity measured for the Cy3 and Cy5 fluors at each array element provides a reliable measure of the relative abundance of the corresponding mRNA in the two cell populations (Fig. 1). Data from the series of seven samples (Fig. 2), consisting of more than 43,000 expression-ratio measurements, were organized into a database to facilitate efficient exploration and analysis of the results. This database is publicly available on the Internet (13).

During exponential growth in glucose-rich medium, the global pattern of gene expression was remarkably stable. Indeed, when gene expression patterns between the first two cell samples (harvested at a 2-hour interval) were compared, mRNA levels differed by a factor of 2 or more for only 19 genes (0.3%), and the largest of these differences was only 2.7-fold (14). However, as glucose was progressively depleted from the growth media during the course of the experiment, a marked change was seen in the global pattern of gene expression. mRNA levels for approximately 710 genes were induced by a factor of at least 2, and the mRNA levels for approximately 1030 genes declined by a factor of at least 2. Messenger RNA levels for 183 genes increased by a factor of at least 4, and mRNA levels for 203 genes diminished by a factor of at least 4. About half of these differentially expressed genes have no currently recognized function and are not yet named. Indeed, more than 400 of the differentially expressed genes have no apparent homology

Department of Biochemistry, Stanford University School of Medicine, Howard Hughes Medical Institute, Stanford, CA 94305–5428, USA.

\*To whom correspondence should be addressed. E-mail: pbrown@crgm.stanford.edu

to any gene whose function is known (15). The responses of these previously uncharacterized genes to the diauxic shift therefore provides the first small clue to their possible roles.

The global view of changes in expression of genes with known functions provides a vivid picture of the way in which the cell adapts to a changing environment. Figure 3 shows a portion of the yeast metabolic pathways involved in carbon and energy metabolism. Mapping the changes we observed in the mRNAs encoding each enzyme onto this framework allowed us to infer the redirection in the flow of metabolites through this system. We observed large inductions of the genes coding for the enzymes aldehyde dehydrogenase (ALD2) and acetyl-coenzyme A (CoA) synthase (ACS1), which function together to convert the products of alcohol dehydrogenase into acetyl-CoA, which in turn is used to fuel the tricarboxylic acid (TCA) cycle and the glyoxylate cycle. The concomitant shutdown of transcription of the genes encoding pyruvate decarboxylase and induction of pyruvate carboxylase rechannels pyruvate away from acetaldehyde, and instead to oxalacetate, where it can serve to supply the TCA cycle and gluconeogenesis. Induction of the pivotal genes *PCK1*, encoding phosphoenolpyruvate carboxykinase, and *FBP1*, encoding fructose 1,6-bisphosphatase, switches the directions of two key irreversible steps in glycolysis, reversing the flow of metabolites along the reversible steps of the glycolytic pathway toward the essential biosynthetic precursor, glucose-6-phosphate. Induction of the genes coding for the trehalose synthase and glycogen synthase complexes promotes channeling of glucose-6-phosphate into these carbohydrate storage pathways.

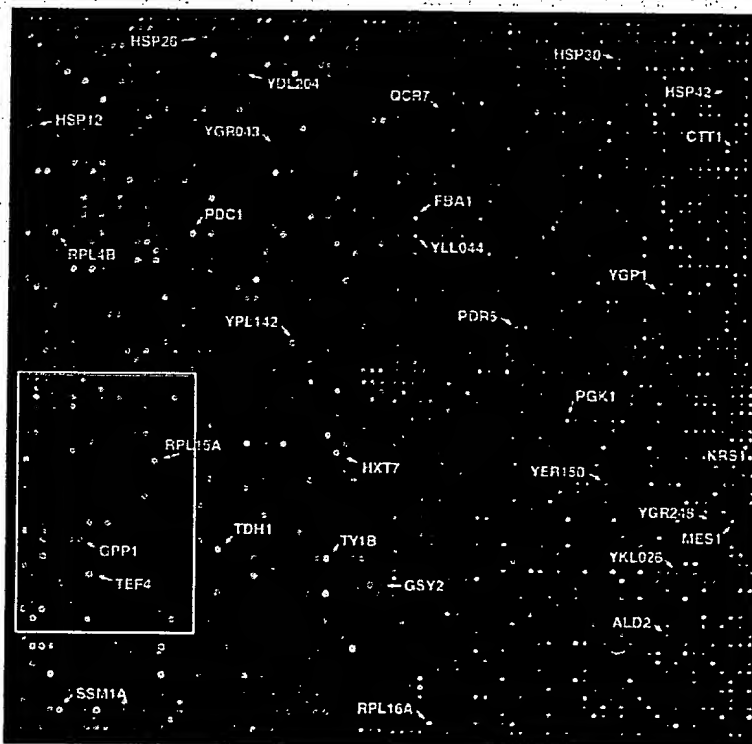
Just as the changes in expression of genes encoding pivotal enzymes can provide insight into metabolic reprogramming, the behavior of large groups of functionally related genes can provide a broad view of the systematic way in which the yeast cell adapts to a changing environment (Fig. 4). Several classes of genes, such as cytochrome c-related genes and those involved in the TCA/glyoxylate cycle and carbohydrate storage, were coordinately induced by glucose exhaustion. In contrast, genes devoted to protein synthesis, including ribosomal proteins, tRNA synthetases, and translation, elongation, and initiation factors, exhibited a coordinated decrease in expression. More than 95% of ribosomal genes showed at least twofold decreases in expression during the diauxic shift (Fig. 4) (13). A noteworthy and illuminating exception was that the

genes encoding mitochondrial ribosomal genes were generally induced rather than repressed after glucose limitation, highlighting the requirement for mitochondrial biogenesis (13). As more is learned about the functions of every gene in the yeast genome, the ability to gain insight into a cell's response to a changing environment through its global gene expression patterns will become increasingly powerful.

Several distinct temporal patterns of expression could be recognized, and sets of genes could be grouped on the basis of the similarities in their expression patterns. The characterized members of each of these groups also shared important similarities in their functions. Moreover, in most cases, common regulatory mechanisms could be inferred for sets of genes with similar expression profiles. For example, seven genes showed a late induction profile, with mRNA levels increasing by more than ninefold at

the last timepoint but less than threefold at the preceding timepoint (Fig. 5B). All of these genes were known to be glucose-repressed, and five of the seven were previously noted to share a common upstream activating sequence (UAS), the carbon source response element (CSRE) (16–20). A search in the promoter regions of the remaining two genes, *ACR1* and *IDP2*, revealed that *ACR1*, a gene essential for ACS1 activity, also possessed a consensus CSRE motif, but interestingly, *IDP2* did not. A search of the entire yeast genome sequence for the consensus CSRE motif revealed only four additional candidate genes, none of which showed a similar induction.

Examples from additional groups of genes that shared expression profiles are illustrated in Fig. 5, C through F. The sequences upstream of the named genes in Fig. 5C all contain stress response elements (STRE), and with the exception



**Fig. 1.** Yeast genome microarray. The actual size of the microarray is 18 mm by 18 mm. The microarray was printed as described (9). This image was obtained with the same fluorescent scanning confocal microscope used to collect all the data we report (49). A fluorescently labeled cDNA probe was prepared from mRNA isolated from cells harvested shortly after inoculation (culture density of  $<5 \times 10^6$  cells/ml and media glucose level of 19 g/liter) by reverse transcription in the presence of Cy3-dUTP. Similarly, a second probe was prepared from mRNA isolated from cells taken from the same culture 9.5 hours later (culture density of  $\sim 2 \times 10^8$  cells/ml, with a glucose level of  $<0.2$  g/liter) by reverse transcription in the presence of Cy5-dUTP. In this image, hybridization of the Cy3-dUTP-labeled cDNA (that is, mRNA expression at the initial timepoint) is represented as a green signal, and hybridization of Cy5-dUTP-labeled cDNA (that is, mRNA expression at 9.5 hours) is represented as a red signal. Thus, genes induced or repressed after the diauxic shift appear in this image as red and green spots, respectively. Genes expressed at roughly equal levels before and after the diauxic shift appear in this image as yellow spots.

of HSP42, have previously been shown to be controlled at least in part by these elements (21–24). Inspection of the sequences upstream of HSP42 and the two uncharacterized genes shown in Fig. 5C, YKL026c, a hypothetical protein with similarity to glutathione peroxidase, and YGR043c, a putative transaldolase, revealed that each of these genes also possess repeated upstream copies of the stress-responsive CCCCT motif. Of the 13 additional genes in the yeast genome that shared this expression profile [including HSP30, ALD2, OM45, and 10 uncharacterized ORFs (25)], nine contained one or more recognizable STRE sites in their upstream regions.

The heterotrimeric transcriptional activator complex HAP2,3,4 has been shown to be responsible for induction of several genes important for respiration (26–28). This complex binds a degenerate consensus sequence known as the CCAAT box (26). Computer analysis, using the consensus sequence TNRYTGGB (29), has suggested that a large number of genes involved in respiration may be specific targets of HAP2,3,4 (30). Indeed, a putative HAP2,3,4 binding site could be found in the sequences upstream of each of the seven cytochrome *c*-related genes that showed the greatest magnitude of induction (Fig. 5D). Of 12 additional cytochrome *c*-related genes that were induced, HAP2,3,4 binding sites were present in all but one. Significantly, we found that transcription of HAP4 itself was induced nearly ninefold concomitant with the diauxic shift.

Control of ribosomal protein biogenesis is mainly exerted at the transcriptional level, through the presence of a common upstream-activating element (UAS<sub>mp</sub>) that is recognized by the Rap1 DNA-binding protein (31, 32). The expression profiles of seven ribosomal proteins are shown in Fig. 5F. A search of the sequences upstream of all seven genes revealed consensus Rap1-binding motifs (33). It has been suggested that declining Rap1 levels in the cell during starvation may be responsible for the decline in ribosomal protein gene expression (34). Indeed, we observed that the abundance of RAP1 mRNA diminished by 4.4-fold, at about the time of glucose exhaustion.

Of the 149 genes that encode known or putative transcription factors, only two, HAP4 and SIP4, were induced by a factor of more than threefold at the diauxic shift. SIP4 encodes a DNA-binding transcriptional activator that has been shown to interact with Snf1, the "master regulator" of glucose repression (35). The eightfold induction of SIP4 upon depletion of glucose strongly suggests a role in the induction of

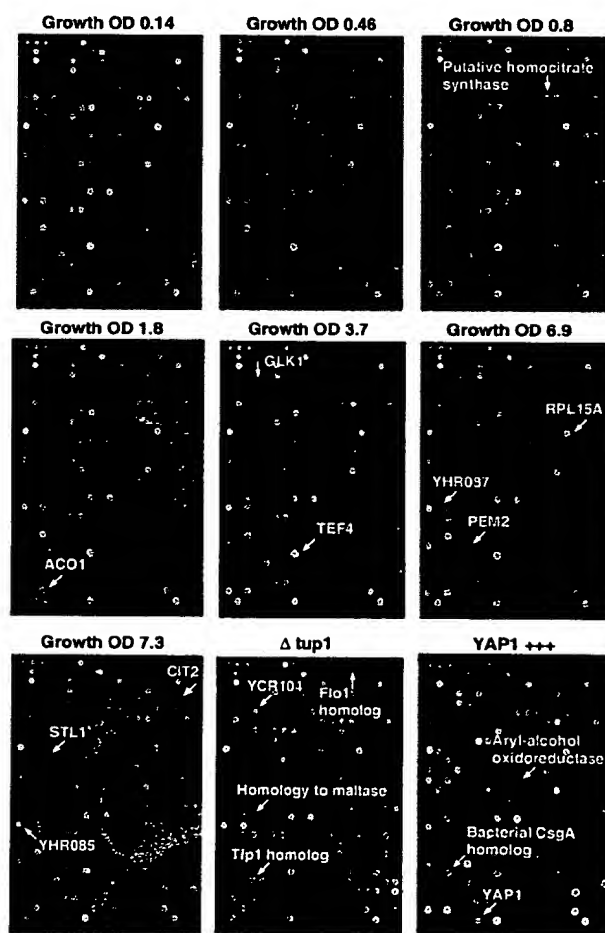
downstream genes at the diauxic shift.

Although most of the transcriptional responses that we observed were not previously known, the responses of many genes during the diauxic shift have been described. Comparison of the results we obtained by DNA microarray hybridization with previously reported results therefore provided a strong test of the sensitivity and accuracy of this approach. The expression patterns we observed for previously characterized genes showed almost perfect concordance with previously published results (36). Moreover, the differential expression measurements obtained by DNA microarray hybridization were reproducible in duplicate experiments. For example, the remarkable changes in gene expression between cells harvested immediately after inoculation and immediately after the diauxic shift (the first and sixth intervals in this time series) were measured in duplicate, independent DNA microarray hybridizations. The correlation coefficient for two complete sets of expression ratio measurements was 0.87, and for more than 95% of the genes, the expres-

sion ratios measured in these duplicate experiments differed by less than a factor of 2. However, in a few cases, there were discrepancies between our results and previous results, pointing to technical limitations that will need to be addressed as DNA microarray technology advances (37, 38). Despite the noted exceptions, the high concordance between the results we obtained in these experiments and those of previous studies provides confidence in the reliability and thoroughness of the survey.

The changes in gene expression during this diauxic shift are complex and involve integration of many kinds of information about the nutritional and metabolic state of the cell. The large number of genes whose expression is altered and the diversity of temporal expression profiles observed in this experiment highlight the challenge of understanding the underlying regulatory mechanisms. One approach to defining the contributions of individual regulatory genes to a complex program of this kind is to use DNA microarrays to identify genes whose expression is affected

**Fig. 2.** The section of the array indicated by the gray box in Fig. 1 is shown for each of the experiments described here. Representative genes are labeled. In each of the arrays used to analyze gene expression during the diauxic shift, red spots represent genes that were induced relative to the initial timepoint, and green spots represent genes that were repressed relative to the initial timepoint. In the arrays used to analyze the effects of the *tup1*Δ mutation and YAP1 overexpression, red spots represent genes whose expression was increased, and green spots represent genes whose expression was decreased by the genetic modification. Note that distinct sets of genes are induced and repressed in the different experiments. The complete images of each of these arrays can be viewed on the Internet (13). Cell density as measured by optical density (OD) at 600 nm was used to measure the growth of the culture.



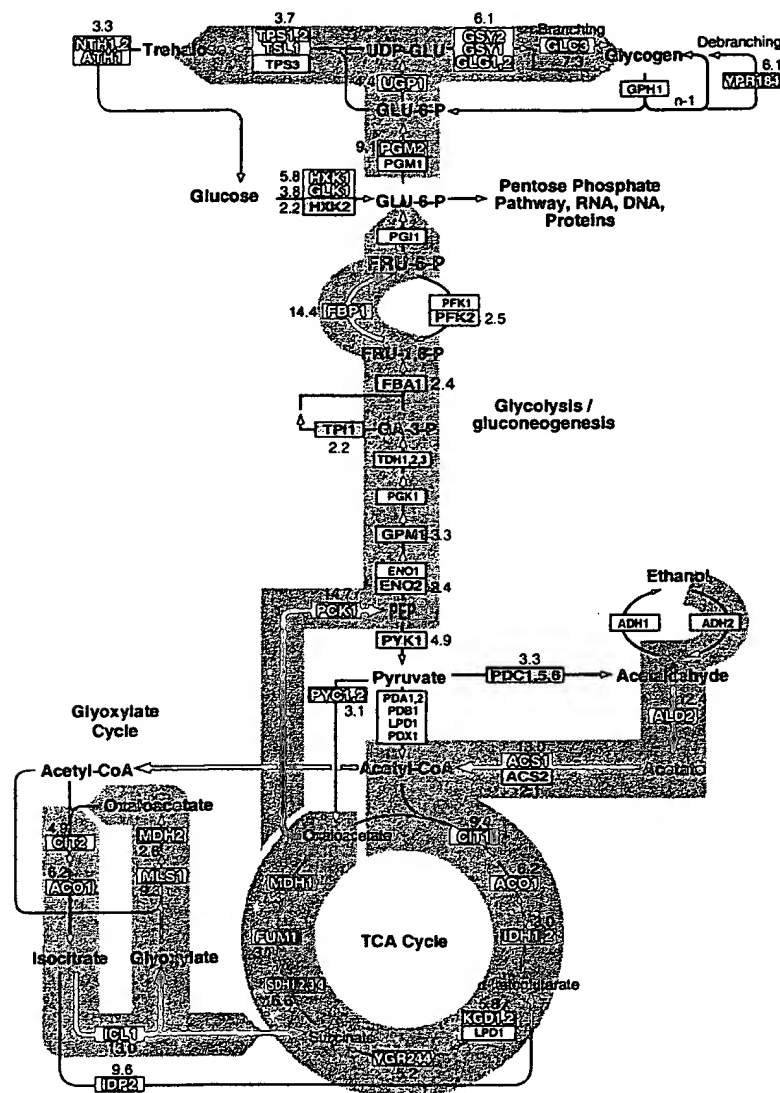
by mutations in each putative regulatory gene. As a test of this strategy, we analyzed the genomewide changes in gene expression that result from deletion of the *TUP1* gene. Transcriptional repression of many genes by glucose requires the DNA-binding repressor

Mig1 and is mediated by recruiting the transcriptional co-repressors Tup1 and Cyc8/Ssn6 (39). Tup1 has also been implicated in repression of oxygen-regulated, mating-type-specific, and DNA-damage-inducible genes (40).

Wild-type yeast cells and cells bearing a deletion of the *TUP1* gene (*tup1Δ*) were grown in parallel cultures in rich medium containing glucose as the carbon source. Messenger RNA was isolated from exponentially growing cells from the two populations and used to prepare cDNA labeled with Cy3 (green) and Cy5 (red), respectively (11). The labeled probes were mixed and simultaneously hybridized to the microarray. Red spots on the microarray therefore represented genes whose transcription was induced in the *tup1Δ* strain, and thus presumably repressed by Tup1 (41). A representative section of the microarray (Fig. 2, bottom middle panel) illustrates that the genes whose expression was affected by the *tup1Δ* mutation, were, in general, distinct from those induced upon glucose exhaustion [complete images of all the arrays shown in Fig. 2 are available on the Internet (13)]. Nevertheless, 34 (10%) of the genes that were induced by a factor of at least 2 after the diauxic shift were similarly induced by deletion of *TUP1*, suggesting that these genes may be subject to *TUP1*-mediated repression by glucose. For example, *SUC2*, the gene encoding invertase, and all five hexose transporter genes that were induced during the course of the diauxic shift were similarly induced, in duplicate experiments, by the deletion of *TUP1*.

The set of genes affected by Tup1 in this experiment also included  $\alpha$ -glucosidases, the mating-type-specific genes *MFA1* and *MFA2*, and the DNA damage-inducible *RNR2* and *RNR4*, as well as genes involved in flocculation and many genes of unknown function. The hybridization signal corresponding to expression of *TUP1* itself was also severely reduced because of the (incomplete) deletion of the transcription unit in the *tup1Δ* strain, providing a positive control in the experiment (42).

Many of the transcriptional targets of Tup1 fell into sets of genes with related biochemical functions. For instance, although only about 3% of all yeast genes appeared to be *TUP1*-repressed by a factor of more than 2 in duplicate experiments under these conditions, 6 of the 13 genes that have been implicated in flocculation (15) showed a reproducible increase in expression of at least twofold when *TUP1* was deleted. Another group of related genes that appeared to be subject to *TUP1* repression encodes the serine-rich cell wall mannoproteins, such as *Tip1* and *Tir1/Srp1* which are induced by cold shock and other stresses (43), and similar, serine-poor proteins, the seripauperins (44). Messenger RNA levels for 23 of the 26 genes in this group were reproducibly elevated by at least 2.5-fold in the *tup1Δ*



**Fig. 3.** Metabolic reprogramming inferred from global analysis of changes in gene expression. Only key metabolic intermediates are identified. The yeast genes encoding the enzymes that catalyze each step in this metabolic circuit are identified by name in the boxes. The genes encoding succinyl-CoA synthase and glycogen-debranching enzyme have not been explicitly identified, but the ORFs YGR244 and YPR184 show significant homology to known succinyl-CoA synthase and glycogen-debranching enzymes, respectively, and are therefore included in the corresponding steps in this figure. Red boxes with white lettering identify genes whose expression increases in the diauxic shift. Green boxes with dark green lettering identify genes whose expression diminishes in the diauxic shift. The magnitude of induction or repression is indicated for these genes. For multimeric enzyme complexes, such as succinate dehydrogenase, the indicated fold-induction represents an unweighted average of all the genes listed in the box. Black and white boxes indicate no significant differential expression (less than twofold). The direction of the arrows connecting reversible enzymatic steps indicate the direction of the flow of metabolites, inferred from the gene expression pattern, after the diauxic shift. Arrows representing steps catalyzed by genes whose expression was strongly induced are highlighted in red. The broad gray arrows represent major increases in the flow of metabolites after the diauxic shift, inferred from the indicated changes in gene expression.

strain, and 18 of these genes were induced by more than sevenfold when *TUP1* was deleted. In contrast, none of 83 genes that could be classified as putative regulators of the cell division cycle were induced more than twofold by deletion of *TUP1*. Thus, despite the diversity of the regulatory systems that employ Tup1, most of the genes that it regulates under these conditions fall into a limited number of distinct functional classes.

Because the microarray allows us to monitor expression of nearly every gene in yeast, we can, in principle, use this approach to identify all the transcriptional targets of a regulatory protein like Tup1. It is important to note, however, that in any single experiment of this kind we can only recognize those target genes that are normally repressed (or induced) under the conditions of the experiment. For instance, the experiment described here analyzed a *MAT*  $\alpha$  strain in which *MFA1* and *MFA2*, the genes encoding the  $\alpha$ -factor mating pheromone precursor, are normally repressed. In the isogenic *tup1* $\Delta$  strain, these genes were inappropriately expressed, reflecting the role that Tup1 plays in their repression. Had we instead carried out this experiment with a *MAT*  $\alpha$  strain (in which expression of *MFA1* and *MFA2* is not repressed), it would not have been possible to conclude anything regarding the role of Tup1 in the repression of these genes. Conversely, we cannot distinguish indirect effects of the chronic absence of Tup1 in the mutant strain from effects directly attributable to its participation in repressing the transcription of a gene.

Another simple route to modulating the activity of a regulatory factor is to overexpress the gene that encodes it. *YAP1* encodes a DNA-binding transcription factor belonging to the b-zip class of DNA-binding proteins. Overexpression of *YAP1* in yeast confers increased resistance to hydrogen peroxide, *o*-phenanthroline, heavy metals, and osmotic stress (45). We analyzed differential gene expression between a wild-type strain bearing a control plasmid and a strain with a plasmid expressing *YAP1* under the control of the strong *GAL1-10* promoter, both grown in galactose (that is, a condition that induces *YAP1* overexpression). Complementary DNA from the control and *YAP1* overexpressing strains, labeled with Cy3 and Cy5, respectively, was prepared from mRNA isolated from the two strains and hybridized to the microarray. Thus, red spots on the array represent genes that were induced in the strain overexpressing *YAP1*.

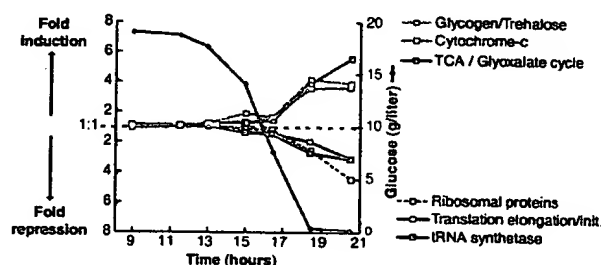
Of the 17 genes whose mRNA levels increased by more than threefold when

*YAP1* was overexpressed in this way, five bear homology to aryl-alcohol oxidoreductases (Fig. 2 and Table 1). An additional four of the genes in this set also belong to the general class of dehydrogenases/oxidoreductases. Very little is known about the role of aryl-alcohol oxidoreductases in *S. cerevisiae*, but these enzymes have been isolated from ligninolytic fungi, in which they participate in coupled redox reactions, oxidizing aromatic, and aliphatic unsaturated alcohols to aldehydes with the production of hydrogen peroxide (46, 47). The fact that a remarkable fraction of the targets identified in this experiment belong to the same small, functional group of oxidoreductases suggests that these genes

might play an important protective role during oxidative stress. Transcription of a small number of genes was reduced in the strain overexpressing *Yap1*. Interestingly, many of these genes encode sugar permeases or enzymes involved in inositol metabolism.

We searched for *Yap1*-binding sites (TTACTAA or TGACTAA) in the sequences upstream of the target genes we identified (48). About two-thirds of the genes that were induced by more than threefold upon *Yap1* overexpression had one or more binding sites within 600 bases upstream of the start codon (Table 1), suggesting that they are directly regulated by *Yap1*. The absence of canonical *Yap1*-bind-

**Fig. 4.** Coordinated regulation of functionally related genes. The curves represent the average induction or repression ratios for all the genes in each indicated group. The total number of genes in each group was as follows: ribosomal proteins, 112; translation elongation and initiation factors, 25; tRNA synthetases (excluding mitochondrial synthetases), 17; glycogen and trehalose synthesis and degradation, 15; cytochrome c oxidase and reductase proteins, 19; and TCA- and glyoxylate-cycle enzymes, 24.



**Table 1.** Genes induced by *YAP1* overexpression. This list includes all the genes for which mRNA levels increased by more than twofold upon *YAP1* overexpression in both of two duplicate experiments, and for which the average increase in mRNA level in the two experiments was greater than threefold (50). Positions of the canonical *Yap1* binding sites upstream of the start codon, when present, and the average fold-increase in mRNA levels measured in the two experiments are indicated.

ORF	Distance of <i>Yap1</i> site from ATG	Gene	Description	Fold-increase
YNL331C			Putative aryl-alcohol reductase	12.9
YKL071W	162-222 (5 sites)	<i>YAP1</i>	Similarity to bacterial <i>csgA</i> protein	10.4
YML007W			Transcriptional activator involved in oxidative stress response	9.8
YFL056C	223, 242		Homology to aryl-alcohol dehydrogenases	9.0
YLL060C	98		Putative glutathione transferase	7.4
YOL165C	266		Putative aryl-alcohol dehydrogenase (NADP+)	7.0
YCR107W			Putative aryl-alcohol reductase	6.5
YML116W	409	<i>ATR1</i>	Aminotriazole and 4-nitroquinoline resistance protein	6.5
YBR008C	142, 167, 364		Homology to benomyl/methotrexate resistance protein	6.1
YCLX08C			Hypothetical protein	6.1
YJR155W			Putative aryl-alcohol dehydrogenase	6.0
YPL171C	148, 212	<i>OYE3</i>	NADPH dehydrogenase (old yellow enzyme), isoform 3	5.8
YLR460C	167, 317		Homology to hypothetical proteins YCR102c and YNL134c	4.7
YKR076W	178		Homology to hypothetical protein YMR251w	4.5
YHR179W	327	<i>OYE2</i>	NAD(P)H oxidoreductase (old yellow enzyme), isoform 1	4.1
YML131W	507		Similarity to <i>A. thaliana</i> zeta-crystallin homolog	3.7
YOL126C		<i>MDH2</i>	Malate dehydrogenase	3.3

ing sites upstream of the others may reflect an ability of Yap1 to bind sites that differ from the canonical binding sites, perhaps in cooperation with other factors, or less likely, may represent an indirect effect of Yap1 overexpression, mediated by one or more intermediary factors. Yap1 sites were found only four times in the corresponding region of an arbitrary set of 30 genes that were not differentially regulated by Yap1.

Use of a DNA microarray to characterize the transcriptional consequences of mutations affecting the activity of regulatory molecules provides a simple and powerful approach to dissection and characterization of regulatory pathways and net-

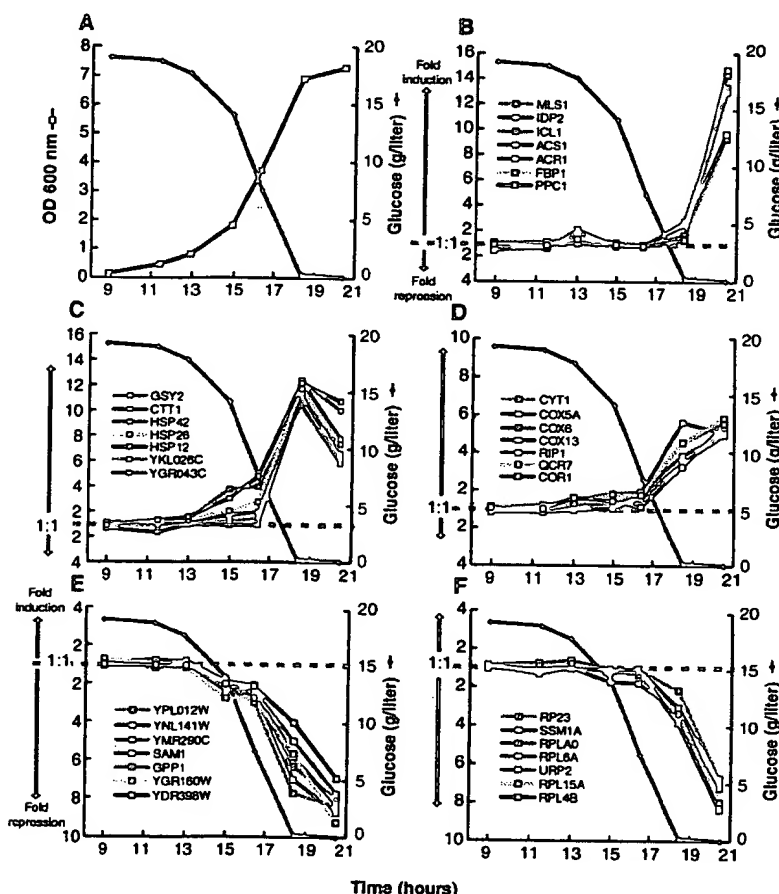
works. This strategy also has an important practical application in drug screening. Mutations in specific genes encoding candidate drug targets can serve as surrogates for the ideal chemical inhibitor or modulator of their activity. DNA microarrays can be used to define the resulting signature pattern of alterations in gene expression, and then subsequently used in an assay to screen for compounds that reproduce the desired signature pattern.

DNA microarrays provide a simple and economical way to explore gene expression patterns on a genomic scale. The hurdles to extending this approach to any other organism are minor. The equipment

required for fabricating and using DNA microarrays (9) consists of components that were chosen for their modest cost and simplicity. It was feasible for a small group to accomplish the amplification of more than 6000 genes in about 4 months and, once the amplified gene sequences were in hand, only 2 days were required to print a set of 110 microarrays of 6400 elements each. Probe preparation, hybridization, and fluorescent imaging are also simple procedures. Even conceptually simple experiments, as we described here, can yield vast amounts of information. The value of the information from each experiment of this kind will progressively increase as more is learned about the functions of each gene and as additional experiments define the global changes in gene expression in diverse other natural processes and genetic perturbations. Perhaps the greatest challenge now is to develop efficient methods for organizing, distributing, interpreting, and extracting insights from the large volumes of data these experiments will provide.

## REFERENCES AND NOTES

1. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* 270, 467 (1995).
2. D. Shalon, S. J. Smith, P. O. Brown, *Genome Res.* 6, 639 (1996).
3. D. Lashkar, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
4. J. DeRisi et al., *Nature Genet.* 14, 457 (1996).
5. D. J. Lockhart et al., *Nature Biotechnol.* 14, 1675 (1996).
6. M. Chee et al., *Science* 274, 610 (1996).
7. M. Johnston and M. Carlson, in *The Molecular Biology of the Yeast Saccharomyces: Gene Expression*, E. W. Jones, J. R. Pringle, J. R. Broach, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992), p. 193.
8. Primers for each known or predicted protein coding sequence were supplied by Research Genetics. PCR was performed with the protocol supplied by Research Genetics, using genomic DNA from yeast strain S288C as a template. Each PCR product was verified by agarose gel electrophoresis and was deemed correct if the lane contained a single band of appropriate mobility. Failures were marked as such in the database. The overall success rate for a single-pass amplification of 6116 ORFs was ~94.5%.
9. Glass slides (Gold Seal) were cleaned for 2 hours in a solution of 2 N NaOH and 70% ethanol. After rinsing in distilled water, the slides were then treated with a 1:5 dilution of poly-L-lysine adhesive solution (Sigma) for 1 hour, and then dried for 5 min at 40°C in a vacuum oven. DNA samples from 100- $\mu$ l PCR reactions were purified by ethanol purification in 96-well microtiter plates. The resulting precipitates were resuspended in 3 $\times$  standard saline citrate (SSC) and transferred to new plates for arraying. A custom-built arraying robot was used to print on a batch of 110 slides. Details of the design of the microarrayer are available at [cmgm.stanford.edu/pbrown](http://cmgm.stanford.edu/pbrown). After printing, the microarrays were rehydrated for 30 s in a humid chamber and then snap-dried for 2 s on a hot plate (100°C). The DNA was then ultraviolet (UV)-crosslinked to the surface by subjecting the slides to 60 mJ of energy (Stratagene Stratelinker). The rest of the poly-L-lysine surface was blocked by a 15-min incubation in a solution of 70 mM succinic anhydride dissolved in a solution consisting of 315 ml of 1-methyl-2-pyrrolidinone (Aldrich) and 35 ml of 1 M boric acid (pH 8.0). Directly after the blocking reac-



**Fig. 5.** Distinct temporal patterns of induction or repression help to group genes that share regulatory properties. (A) Temporal profile of the cell density, as measured by OD at 600 nm and glucose concentration in the media. (B) Seven genes exhibited a strong induction (greater than ninefold) only at the last timepoint (20.5 hours). With the exception of *IDP2*, each of these genes has a CSRE UAS. There were no additional genes observed to match this profile. (C) Seven members of a class of genes marked by early induction with a peak in mRNA levels at 18.5 hours. Each of these genes contains STRE motif repeats in their upstream promoter regions. (D) Cytochrome c oxidase and ubiquinol cytochrome c reductase genes. Marked by an induction coincident with the diauxic shift, each of these genes contains a consensus binding motif for the HAP2,3,4 protein complex. At least 17 genes shared a similar expression profile. (E) *SAM1*, *GPP1*, and several genes of unknown function are repressed before the diauxic shift, and continue to be repressed upon entry into stationary phase. (F) Ribosomal protein genes comprise a large class of genes that are repressed upon depletion of glucose. Each of the genes profiled here contains one or more RAP1-binding motifs upstream of its promoter. RAP1 is a transcriptional regulator of most ribosomal proteins.



- tion, the bound DNA was denatured by a 2-min incubation in distilled water at  $-95^{\circ}\text{C}$ . The slides were then transferred into a bath of 100% ethanol at room temperature, rinsed, and then spun dry in a clinical centrifuge. Slides were stored in a closed box at room temperature until used.
10. YPD medium (8 liters), in a 10-liter fermentation vessel, was inoculated with 2 ml of a fresh overnight culture of yeast strain DBY7286 (MATa, ura3, GAL2). The fermentor was maintained at  $30^{\circ}\text{C}$  with constant agitation and aeration. The glucose content of the media was measured with a UV test kit (Boehringer Mannheim, catalog number 716251). Cell density was measured by OD at 600-nm wavelength. Aliquots of culture were rapidly withdrawn from the fermentation vessel by peristaltic pump, spun down at room temperature, and then flash frozen with liquid nitrogen. Frozen cells were stored at  $-80^{\circ}\text{C}$ .
  11. Cy3-dUTP or Cy5-dUTP (Amersham) was incorporated during reverse transcription of 1.25  $\mu\text{g}$  of polyadenylated [poly(A) $^{+}$ ] RNA, primed by a dT(16) oligomer. This mixture was heated to  $70^{\circ}\text{C}$  for 10 min, and then transferred to ice. A premixed solution, consisting of 200 U Superscript II (Gibco), buffer, deoxyribonucleoside triphosphates, and fluorescent nucleotides, was added to the RNA. Nucleotides were used at these final concentrations: 500  $\mu\text{M}$  for dATP, dCTP, and dGTP and 200  $\mu\text{M}$  for dTTP. Cy3-dUTP and Cy5-dUTP were used at a final concentration of 100  $\mu\text{M}$ . The reaction was then incubated at  $42^{\circ}\text{C}$  for 2 hours. Unincorporated fluorescent nucleotides were removed by first diluting the reaction mixture with 470  $\mu\text{l}$  of 10 mM tris-HCl (pH 8.0)/1 mM EDTA and then subsequently concentrating the mix to  $\sim 5 \mu\text{l}$ , using Centricon-30 microconcentrators (Amicon).
  12. Purified, labeled cDNA was resuspended in 11  $\mu\text{l}$  of  $3.5\times$  SSC containing 10  $\mu\text{g}$  poly(dA) and 0.3  $\mu\text{l}$  of 10% SDS. Before hybridization, the solution was boiled for 2 min and then allowed to cool to room temperature. The solution was applied to the microarray under a cover slip, and the slide was placed in a custom hybridization chamber which was subsequently incubated for  $\sim 8$  to 12 hours in a water bath at  $62^{\circ}\text{C}$ . Before scanning, slides were washed in  $2\times$  SSC, 0.2% SDS for 5 min, and then  $0.05\times$  SSC for 1 min. Slides were dried before scanning by centrifugation at 500 rpm in a Beckman CS-6R centrifuge.
  13. The complete data set is available on the Internet at [cmgm.stanford.edu/pbrown/explore/index.html](http://cmgm.stanford.edu/pbrown/explore/index.html)
  14. For 95% of all the genes analyzed, the mRNA levels measured in cells harvested at the first and second interval after inoculation differed by a factor of less than 1.5. The correlation coefficient for the comparison between mRNA levels measured for each gene in these two different mRNA samples was 0.98. When duplicate mRNA preparations from the same cell sample were compared in the same way, the correlation coefficient between the expression levels measured for the two samples by comparative hybridization was 0.99.
  15. The numbers and identities of known and putative genes, and their homologies to other genes, were gathered from the following public databases: Saccharomyces Genome Database ([genome-www.stanford.edu](http://genome-www.stanford.edu)), Yeast Protein Database ([quest7.proteome.com](http://quest7.proteome.com)), and Munich Information Centre for Protein Sequences ([speedy.mips.biochem.mpg.de/mips/yeast/index.html](http://speedy.mips.biochem.mpg.de/mips/yeast/index.html)).
  16. A. Scholer and H. J. Schuller, *Mol. Cell. Biol.* 14, 3613 (1994).
  17. S. Kratzer and H. J. Schuller, *Gene* 161, 75 (1995).
  18. R. J. Haselbeck and H. L. McAlister, *J. Biol. Chem.* 268, 12116 (1993).
  19. M. Fernandez, E. Fernandez, R. Rodicio, *Mol. Gen. Genet.* 242, 727 (1994).
  20. A. Hartig et al., *Nucleic Acids Res.* 20, 5677 (1992).
  21. P. M. Martinez et al., *EMBO J.* 15, 2227 (1996).
  22. J. C. Varela, U. M. Prækel, P. A. Meacock, R. J. Planta, W. H. Mager, *Mol. Cell. Biol.* 15, 6232 (1995).
  23. H. Ruis and C. Schuller, *Bioessays* 17, 959 (1995).
  24. J. L. Parrou, M. A. Teste, J. Francois, *Microbiology* 143, 1891 (1997).
  25. This expression profile was defined as having an induction of greater than 10-fold at 18.5 hours and less than 11-fold at 20.5 hours.
  26. S. L. Forsburg and L. Guarente, *Genes Dev.* 3, 1166 (1989).
  27. J. T. Olesen and L. Guarente, *ibid.* 4, 1714 (1990).
  28. M. Rosenkrantz, C. S. Kell, E. A. Pennell, L. J. Deverish, *Mol. Microbiol.* 13, 119 (1994).
  29. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. The nucleotide codes are as follows: B-C, G, or T; N-G, A, T, or C; R-A or G; and Y-C or T.
  30. C. Fondrat and A. Kalogeropoulos, *Comput. Appl. Biosci.* 12, 363 (1996).
  31. D. Shore, *Trends Genet.* 10, 408 (1994).
  32. R. J. Planta and H. A. Reue, *ibid.* 4, 64 (1988).
  33. The degenerate consensus sequence VYCYRNNC-MNH was used to search for potential RAP1-binding sites. The exact consensus, as defined by (30), is WACAYCORTACATYWW, with up to three differences allowed.
  34. S. F. Neuman, S. Bhattacharya, J. R. Broach, *Mol. Cell. Biol.* 15, 3187 (1995).
  35. P. Lesage, X. Yang, M. Carlson, *ibid.* 16, 1921 (1996).
  36. For example, we observed large inductions of the genes coding for PCK1, FBP1 [Z. Yin et al., *Mol. Microbiol.* 20, 751 (1996)], the central glyoxylate cycle gene ICL1 [A. Scholer and H. J. Schuller, *Curr. Genet.* 23, 375 (1993)], and the "aerobic" isoform of acetyl-CoA synthase, ACS1 [M. A. van den Berg et al., *J. Biol. Chem.* 271, 28953 (1996)], with concomitant down-regulation of the glycolytic-specific genes PYK1 and PFK2 [P. A. Moore et al., *Mol. Cell. Biol.* 11, 5330 (1991)]. Other genes not directly involved in carbon metabolism but known to be induced upon nutrient limitation include genes encoding cytosolic catalase T CTT1 [P. H. Bissinger et al., *ibid.* 9, 1309 (1989)] and several genes encoding small heat-shock proteins, such as HSP12, HSP26, and HSP42 [I. Farkas et al., *J. Biol. Chem.* 266, 15602 (1991); U. M. Prækel and P. A. Meacock, *Mol. Gen. Genet.* 223, 97 (1990); D. Wotton et al., *J. Biol. Chem.* 271, 2717 (1996)].
  37. The levels of induction we measured for genes that were expressed at very low levels in the uninduced state (notably, FBP1 and PCK1) were generally lower than those previously reported. This discrepancy was likely due to the conservative background subtraction method we used, which generally resulted in overestimation of very low expression levels (46).
  38. Cross-hybridization of highly related sequences can also occasionally obscure changes in gene expression, an important concern where members of gene families are functionally specialized and differentially regulated. The major alcohol dehydrogenase genes, ADH1 and ADH2, share 88% nucleotide identity. Reciprocal regulation of these genes is an important feature of the diauxic shift, but was not observed in this experiment, presumably because of cross-hybridization of the fluorescent cDNAs representing these two genes. Nevertheless, we were able to detect differential expression of closely related isoforms of other enzymes, such as HXK1/HXK2 (77% identical) [P. Herrero et al., *Yeast* 11, 137 (1995)], MLS1/DAL7 (73% identical) (20), and PGM1/PGM2 (72% identical) [D. Oh, J. E. Hopper, *Mol. Cell. Biol.* 10, 1415 (1990)], in accord with previous studies. Use in the microarray of deliberately selected DNA sequences corresponding to the most divergent segments of homologous genes, in lieu of the complete gene sequences, should relieve this problem in many cases.
  39. F. E. Williams, U. Varanasi, R. J. Trumbly, *Mol. Cell. Biol.* 11, 3307 (1991).
  40. D. Tzamaras and K. Struhl, *Nature* 369, 758 (1994).
  41. Differences in mRNA levels between the *tup1 $\Delta$*  and wild-type strain were measured in two independent experiments. The correlation coefficient between the complete sets of expression ratios measured in these duplicate experiments was 0.83. The concordance between the sets of genes that appeared to be induced was very high between the two experiments. When only the 355 genes that showed at least a twofold increase in mRNA in the *tup1 $\Delta$*  strain in either of the duplicate experiments were compared, the correlation coefficient was 0.82.
  42. The *tup1 $\Delta$*  mutation consists of an insertion of the LEU2 coding sequence, including a stop codon, between the ATG of TUP1 and an Eco RI site 124 base pairs before the stop codon of the TUP1 gene.
  43. L. R. Kowalski, K. Kondo, M. Inouye, *Mol. Microbiol.* 15, 341 (1995).
  44. M. Viswanathan, G. Muthukumar, Y. S. Cong, J. Lenard, *Gene* 148, 149 (1994).
  45. D. Hirata, K. Yano, T. Miyakawa, *Mol. Gen. Genet.* 242, 250 (1994).
  46. A. Gutierrez, L. Caramelo, A. Prieto, M. J. Martinez, A. T. Martinez, *Appl. Environ. Microbiol.* 60, 1783 (1994).
  47. A. Muheim et al., *Eur. J. Biochem.* 195, 369 (1991).
  48. J. A. Wemmie, M. S. Szczypka, D. J. Thiele, W. S. Moye-Rowley, *J. Biol. Chem.* 269, 32592 (1994).
  49. Microarrays were scanned using a custom-built scanning laser microscope built by S. Smith with software written by N. Ziv. Details concerning scanner design and construction are available at [cmgm.stanford.edu/pbrown](http://cmgm.stanford.edu/pbrown). Images were scanned at a resolution of 20  $\mu\text{m}$  per pixel. A separate scan, using the appropriate excitation line, was done for each of the two fluorophores used. During the scanning process, the ratio between the signals in the two channels was calculated for several array elements containing total genomic DNA. To normalize the two channels with respect to overall intensity, we then adjusted photomultiplier and laser power settings such that the signal ratio at these elements was as close to 1.0 as possible. The combined images were analyzed with custom-written software. A bounding box, fitted to the size of the DNA spots in each quadrant, was placed over each array element. The average fluorescent intensity was calculated by summing the intensities of each pixel present in a bounding box, and then dividing by the total number of pixels. Local area background was calculated for each array element by determining the average fluorescent intensity for the lower 20% of pixel intensities. Although this method tends to underestimate the background, causing an underestimation of extreme ratios, it produces a very consistent and noise-tolerant approximation. Although the analog-to-digital board used for data collection possesses a wide dynamic range (12 bits), several signals were saturated (greater than the maximum signal intensity allowed) at the chosen settings. Therefore, extreme ratios at bright elements are generally underestimated. A signal was deemed significant if the average intensity after background subtraction was at least 2.5-fold higher than the standard deviation in the background measurements for all elements on the array.
  50. In addition to the 17 genes shown in Table 1, three additional genes were induced by an average of more than threefold in the duplicate experiments, but in one of the two experiments, the induction was less than twofold (range 1.6- to 1.9-fold).
  51. We thank H. Bennett, P. Spellman, J. Ravetto, M. Eisen, R. Pillai, B. Dunn, T. Ferea, and other members of the Brown lab for their assistance and helpful advice. We also thank S. Friend, D. Botstein, S. Smith, J. Hudson, and D. Dolginow for advice, support, and encouragement; K. Struhl and S. Chatterjee for the *Tup1* deletion strain; L. Fernandes for helpful advice on Yap1; and S. Klapholz and the reviewers for many helpful comments on the manuscript. Supported by a grant from the National Human Genome Research Institute (NHGRI) (HG00450), and by the Howard Hughes Medical Institute (HHMI). J.D.R. was supported by the HHMI and the NHGRI. V.R. was supported in part by an Institutional Training Grant in Genome Science (T32 HG00044) from the NHGRI. P.O.B. is an associate investigator of the HHMI.

5 September 1997; accepted 22 September 1997